

# THE EVOLUTION OF THE GALAXY REST-FRAME ULTRAVIOLET LUMINOSITY FUNCTION OVER THE FIRST TWO BILLION YEARS

STEVEN L. FINKELSTEIN<sup>1</sup>, RUSSELL E. RYAN, JR.<sup>2</sup>, CASEY PAPOVICH<sup>3</sup>, MARK DICKINSON<sup>4</sup>, MIMI SONG<sup>1</sup>, RACHEL S. SOMERVILLE<sup>5</sup>,  
HENRY C. FERGUSON<sup>2</sup>, BRETT SALMON<sup>3</sup>, MAURO GIAVALISCO<sup>6</sup>, ANTON M. KOEKEMOER<sup>2</sup>, MATTHEW L. N. ASHBY<sup>7</sup>,  
PETER BEHROOZI<sup>2</sup>, MARCO CASTELLANO<sup>8</sup>, JAMES S. DUNLOP<sup>9</sup>, SANDY M. FABER<sup>10</sup>, GIOVANNI G. FAZIO<sup>7</sup>, ADRIANO FONTANA<sup>8</sup>,  
NORMAN A. GROGIN<sup>2</sup>, NIMISH HATHI<sup>11</sup>, JASON JAACKS<sup>1</sup>, DALE D. KOCEVSKI<sup>12</sup>, RACHAEL LIVERMORE<sup>1</sup>, ROSS J. MCLURE<sup>9</sup>,  
EMILIANO MERLIN<sup>8</sup>, BAHRAM MOBASHER<sup>13</sup>, JEFFREY A. NEWMAN<sup>14</sup>, MARC RAFELSKI<sup>15</sup>, VITHAL TILVI<sup>3</sup>, AND S. P. WILLNER<sup>7</sup>

<sup>1</sup>Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA; [stevenf@astro.as.utexas.edu](mailto:stevenf@astro.as.utexas.edu)

<sup>2</sup>Space Telescope Science Institute, Baltimore, MD 21218, USA

<sup>3</sup>George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA

<sup>4</sup>National Optical Astronomy Observatory, Tucson, AZ 85719, USA

<sup>5</sup>Department of Physics & Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, NJ 08854, USA

<sup>6</sup>Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA

<sup>7</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>8</sup>INAF—Osservatorio Astronomico di Roma, via Frascati 33, I-00040 Monteporzio, Italy

<sup>9</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, UK

<sup>10</sup>University of California Observatories/Lick Observatory, University of California, Santa Cruz, CA, 95064, USA

<sup>11</sup>Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, F-13388, Marseille, France

<sup>12</sup>Department of Physics and Astronomy, Colby College, Waterville, ME 04901, USA

<sup>13</sup>Department of Physics and Astronomy, University of California Riverside, Riverside, CA 92512, USA

<sup>14</sup>Department of Physics and Astronomy and Pitt-PACC, University of Pittsburgh, Pittsburgh, PA 15260, USA

<sup>15</sup>NASA Postdoctoral Program Fellow, Goddard Space Flight Center, Code 665, Greenbelt, MD 20771, USA

Received 2014 October 20; accepted 2015 June 28; published 2015 September 1

## ABSTRACT

We present a robust measurement and analysis of the rest-frame ultraviolet (UV) luminosity functions at  $z = 4\text{--}8$ . We use deep *Hubble Space Telescope* imaging over the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey/GOODS fields, the Hubble Ultra Deep Field, and the Hubble Frontier Field deep parallel observations near the Abell 2744 and MACS J0416.1-2403 clusters. The combination of these surveys provides an effective volume of  $0.6\text{--}1.2 \times 10^6$  Mpc<sup>3</sup> over this epoch, allowing us to perform a robust search for faint ( $M_{UV} = -18$ ) and bright ( $M_{UV} < -21$ ) high-redshift galaxies. We select candidate galaxies using a well-tested photometric redshift technique with careful screening of contaminants, finding a sample of 7446 candidate galaxies at  $3.5 < z < 8.5$ , with  $>1000$  galaxies at  $z \approx 6\text{--}8$ . We measure both a stepwise luminosity function for candidate galaxies in our redshift samples, and a Schechter function, using a Markov Chain Monte Carlo analysis to measure robust uncertainties. At the faint end, our UV luminosity functions agree with previous studies, yet we find a higher abundance of UV-bright candidate galaxies at  $z \geq 6$ . Our best-fit value of the characteristic magnitude  $M_{UV}^*$  is consistent with  $-21$  at  $z \geq 5$ , which is different than that inferred based on previous trends at lower redshift, and brighter at  $\sim 2\sigma$  significance than previous measures at  $z = 6$  and  $7$ . At  $z = 8$ , a single power law provides an equally good fit to the UV luminosity function, while at  $z = 6$  and  $7$  an exponential cutoff at the bright end is moderately preferred. We compare our luminosity functions to semi-analytical models, and find that the lack of evolution in  $M_{UV}^*$  is consistent with models where the impact of dust attenuation on the bright end of the luminosity function decreases at higher redshift, although a decreasing impact of feedback may also be possible. We measure the evolution of the cosmic star-formation rate (SFR) density by integrating our observed luminosity functions to  $M_{UV} = -17$ , correcting for dust attenuation, and find that the SFR density declines proportionally to  $(1+z)^{-4.3 \pm 0.5}$  at  $z > 4$ , which is consistent with observations at  $z \geq 9$ . Our observed luminosity functions are consistent with a reionization history that starts at  $z \gtrsim 10$ , completes at  $z > 6$ , and reaches a midpoint ( $x_{HII} = 0.5$ ) at  $6.7 < z < 9.4$ . Finally, using a constant cumulative number density selection and an empirically derived rising star-formation history, our observations predict that the abundance of bright  $z = 9$  galaxies is likely higher than previous constraints, although consistent with recent estimates of bright  $z \sim 10$  galaxies.

*Key words:* early universe – galaxies: evolution – galaxies: formation – galaxies: high-redshift – ultraviolet: galaxies

*Supporting material:* machine-readable table

## 1. INTRODUCTION

The past half-decade has seen a remarkable increase in our understanding of galaxy evolution over the first billion years after the Big Bang, primarily due to the updated near-infrared capabilities of the *Hubble Space Telescope* (*HST*). Robust

galaxy samples at  $z > 6$  now include more than 1000 objects (e.g., Bouwens et al. 2010a; Bunker et al. 2010; Finkelstein et al. 2010, 2012b; McLure et al. 2010; Oesch et al. 2010b, 2012; Yan et al. 2012), with a few candidate galaxies having likely redshifts as high as 10 (e.g., Bouwens et al. 2011a, 2015; Coe et al. 2013; Ellis et al. 2013; McLure et al. 2013; Oesch

et al. 2013, 2014). These galaxies are selected photometrically, primarily based on a sharp break at rest-frame 1216 Å due to absorption by intervening neutral hydrogen in the intergalactic medium (IGM).

Studies of galaxies at  $z > 6$  have revealed a number of interesting results. Galaxies at  $6 < z < 8$  appear to have bluer rest-frame ultraviolet (UV) colors than at lower redshift, likely due to a decrease in dust attenuation, although the brightest/most massive galaxies do appear to have comparable levels of dust attenuation at  $z = 4\text{--}7$  (e.g., Stanway et al. 2005; Bouwens et al. 2010b; Finkelstein et al. 2010, 2012b; Wilkins et al. 2011; Dunlop et al. 2013; Bouwens et al. 2014). Lower mass galaxies have colors consistent with stellar populations harboring significant metal content (though likely sub-solar), and therefore the currently detectable populations of galaxies are not dominated by the primordial first generation of stars (e.g., Dunlop et al. 2012, 2013; Finkelstein et al. 2012b). The structures of these galaxies are resolvable, although they show small sizes with half-light radii  $\leq 1$  kpc, which is consistent with the evolution previously detected at lower redshifts (e.g., Ferguson et al. 2004; Oesch et al. 2010a; Ono et al. 2013). Finally, the abundance of high-redshift star-forming galaxies may account for the necessary photons to sustain an ionized IGM by  $z \sim 6$ , and perhaps as high as  $z = 7\text{--}8$ , if one assumes that galaxies at least 5 mag below the detection limit of *HST* exist (e.g., Finkelstein et al. 2012a; Robertson et al. 2013); however, the unknown ionizing photon escape fraction is a major systematic uncertainty.

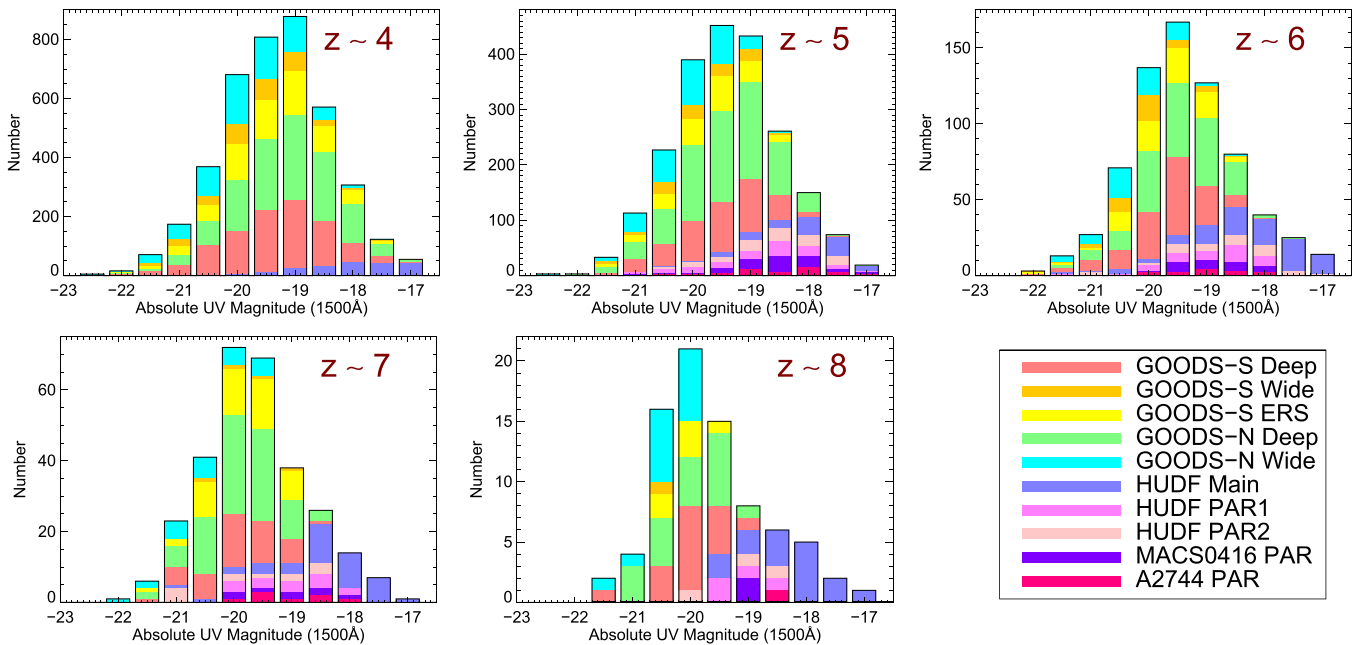
One of the key measurements is the galaxy rest-frame UV luminosity function (hereafter, the luminosity function), because it is one of the most useful tools to study the evolution of a galaxy population. This measure encapsulates the relative abundances of galaxies over a wide dynamic range in luminosity. As the UV light probes recent star-formation activity, the integral of the rest-UV luminosity function provides an estimate of the cosmic star-formation rate (SFR) density (e.g., Madau et al. 1996; Bouwens et al. 2012; Madau & Dickinson 2014), although this measurement is reliant on dust corrections. The luminosity function is typically parameterized with a Schechter (1976) function with a power law slope at faint luminosities, and an exponentially declining form at the bright end. Previous studies have compared the shape of the luminosity function to the underlying dark-matter halo mass function, and found that the luminosity function at  $z \leq 6$ , when normalized to the halo mass function at the characteristic magnitude  $M_{\text{UV}}^*$ , lies below the halo mass function at both bright and faint luminosities. This is generally assumed to be due to feedback: dominated by accreting supermassive black holes at the bright end (active galactic nuclei; AGN), and by supernova or radiative-driven winds at the faint end (e.g., Somerville et al. 2008). Dust extinction can also play a role, particularly if the level of attenuation is dependent on a galaxy’s stellar mass or UV luminosity (e.g., Finkelstein et al. 2012b; Bouwens et al. 2014). Although luminous AGN are present at  $z = 6$  (e.g., Fan et al. 2006), they are exceedingly rare, and to date only a single quasar has been observed at  $z \geq 7$  (Mortlock et al. 2011). Therefore, one may expect the degree of the exponential decline at the bright end to become weaker with increasing redshift. In addition, robustly quantifying the bright end of the luminosity function can allow us to gain physical insight into how these distant galaxies turn their gas into stars. The star-formation timescale is a significant fraction

of the age of the universe, therefore enough time has not yet elapsed for feedback to bring these galaxies into equilibrium. A change in the star-formation timescale is therefore more readily apparent in the shape of the bright end of the luminosity function (e.g., Somerville et al. 2012).

Thanks to the combination of observations from *GALEX* and *HST*, estimates of the UV luminosity function now exist from  $z < 1$  (Arnouts et al. 2005; Cucciati et al. 2012) out to  $z \geq 8$  (e.g., Bouwens et al. 2007, 2011b; McLure et al. 2009; Oesch et al. 2012, 2013; Lorenzoni et al. 2013). Earlier works have concluded that  $M_{\text{UV}}^*$  declines from around  $-21$  at  $z = 3$  to fainter than  $-20$  at  $z = 8$ , with the faint-end slope  $\alpha$  becoming steeper over this same redshift range (e.g., Bouwens et al. 2007, 2011b; Reddy & Steidel 2009; McLure et al. 2013; Schenker et al. 2013). However, in order to adequately quantify the amplitude and form of the bright end, large volumes need to be probed because bright sources are relatively rare. This has been accomplished via a combination of ground- and space-based surveys at  $z \leq 6$ , with a variety of studies showing conclusively that a single power law does not fit the data, and that some sort of cutoff is needed at the bright end (e.g., Arnouts et al. 2005; Bouwens et al. 2007; McLure et al. 2009; Reddy & Steidel 2009). Although previous luminosity functions have been published at  $z \geq 6$ , these space-based studies were based on small volumes (e.g., Bouwens et al. 2011b), and thus, while they can somewhat constrain the faint-end slope, they do not have the capability to constrain the bright end.

Recent studies are starting to make progress at the bright end. Finkelstein et al. (2013), while selecting galaxies for spectroscopic follow up in the GOODS-N field, found an overabundance of bright galaxies at  $z = 7$ . Ono et al. (2012) found a similar result, with their discovery of the  $M_{\text{UV}} = -21.8$  galaxy GN-108036 at  $z = 7.2$  in GOODS-N. Likewise, Hathi et al. (2012) found two bright  $z > 6.5$  candidate galaxies in a ground-based near-infrared survey of GOODS-N. Thus, it appears that the abundance of galaxies at the bright end of the luminosity function may not be decreasing toward higher redshift as previously thought. Although these studies were based in a single field, further evidence comes from Bowler et al. (2014), who used new deep ground-based near-infrared imaging from the UltraVISTA survey (McCracken et al. 2012) to discover 34 luminous  $z \sim 7$  galaxy candidates over  $1.65 \text{ deg}^2$ . They combined these galaxies with the results from McLure et al. (2013)—which included deep and wide *HST* imaging over  $300 \text{ arcmin}^2$  in the GOODS-S, UDS, and Hubble Ultra Deep Field (HUDF) fields—to analyze the rest-frame UV luminosity function at  $z = 7$ . They concluded that they did see evidence for a drop-off in the luminosity function at the bright end; however, the drop-off was less steep than that predicted by a Schechter function, leading them to postulate that the  $z = 7$  luminosity function has the shape of a double power law, perhaps similar to that of the possible form of far-infrared luminosity functions (Sanders et al. 2003; Casey et al. 2014a).

In this study, we measure the rest-frame UV luminosity function at  $4 < z < 8$  with solely space-based data, using the largest *HST* project ever: the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; PIs Faber & Ferguson). The large area observed by CANDELS allows us to probe large volumes of the distant universe for rare, bright galaxies. With these data, we investigate the form of the bright end of the luminosity function and the implications on galaxy



**Figure 1.** Absolute magnitude distribution of all candidate galaxies in our redshift samples. The shaded color denotes which subfield a given galaxy was detected in. This figure demonstrates that while the HUDF is useful for finding the faintest galaxies, CANDELS imaging is necessary to discover much larger numbers, as well as to probe a large dynamic range in luminosity.

evolution. In addition to the deep data in the HUDF, we use the CANDELS data in the GOODS-S and GOODS-N fields, which have not only deeper near-infrared imaging, but also imaging in more optical and near-infrared filters than the other three CANDELS fields (UDS, EGS, and COSMOS). We also include in our analysis the parallel fields from the first year data set of the Hubble Frontier Fields (HFF), which are near the Abell 2744 and MACS J0416.1-2403 galaxy clusters. The combination of these data allows us to select a large sample of nearly 7500 galaxies over a wide dynamic range in UV luminosity at  $z = 4-8$  (Figure 1).

This paper is organized as follows. In Section 2 we discuss the imaging data used and the catalog construction, and in Section 3 we present our sample selection via photometric redshifts, and estimates of the contamination. In Section 4 we highlight our completeness simulations, and in Section 5 we discuss the construction of the rest-UV luminosity function at  $z = 4, 5, 6, 7$ , and 8. In Section 6 we discuss the implications of our luminosity function results, while in Section 7 we compare our results to semi-analytical models. In Section 8 we present our measurements of the cosmic SFR density, and in Section 9 we discuss the implications for galaxies at higher redshifts. Our conclusions are presented in Section 10. Throughout this paper we assume a  $\Lambda$ CDM cosmology (Komatsu et al. 2011), with  $H_0 = 70.2 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_M = 0.275$ ,  $\Omega_\Lambda = 0.725$ , and  $\sigma_8 = 0.816$ . All magnitudes given are in the AB system (Oke & Gunn 1983). All error bars shown in the figures represent  $1\sigma$  uncertainties (or central 68% confidence ranges), unless otherwise stated.

## 2. OBSERVATIONS AND PHOTOMETRY

### 2.1. Imaging Data

Studying galaxies in the early universe requires extremely deep imaging, necessitating space-based data. Additionally, we need to combine deep studies over small areas with larger-area

surveys with shallower limiting magnitudes to probe a large dynamic range in luminosities. Our study used imaging data from a number of surveys covering both the northern and southern fields from the Great Observatories Origins Deep Survey (Giavalisco et al. 2004), with both the *HST* and the *Spitzer Space Telescope*.

The deepest imaging comes from three surveys of the HUDF: the original HUDF survey, which obtained optical imaging with the Advanced Camera for Surveys (ACS; Beckwith et al. 2006), and the more recent HUDF09 (PI Illingworth; e.g., Bouwens et al. 2010a; Oesch et al. 2010b) and UDF12 surveys (PI Ellis; Ellis et al. 2013; Koekemoer et al. 2013), which obtained near-infrared imaging with the Wide Field Camera 3 (WFC3). The full *HST* data set over the HUDF comprises imaging in eight bands, F435W, F606W, F775W, and F850LP with ACS; and F105W, F125W, F140W, and F160W with WFC3 (hereafter  $B_{435}$ ,  $V_{606}$ ,  $i_{775}$ ,  $z_{850}$ ,  $Y_{105}$ ,  $J_{125}$ ,  $JH_{140}$ , and  $H_{160}$ , respectively), which cover an area of  $\sim 5 \text{ arcmin}^2$ . The HUDF09 survey also obtained deep WFC3 imaging over two similarly sized flanking fields, first observed with ACS in the UDF05 survey (PI Stiavelli; Oesch et al. 2007), referred to as the HUDF09-01 and HUDF09-02 fields (Bouwens et al. 2011b). These fields each have imaging in the  $V_{606}$ ,  $i_{775}$ ,  $z_{850}$ ,  $Y_{105}$ ,  $J_{125}$ , and  $H_{160}$  bands.

The majority of our candidate galaxy sample comes from CANDELS (PIs Faber and Ferguson; Grogin et al. 2011; Koekemoer et al. 2011). The largest *HST* project ever, CANDELS comprises 902 orbits over five extragalactic deep fields, including the two GOODS fields (Giavalisco et al. 2004). CANDELS, which finished in 2013 August, is composed of a deep and a wide survey. The deep survey covers the central  $\sim 50\%$  of each of the two GOODS fields, while the wide survey covers the remainder of the GOODS-N field, as well as the southern  $\sim 25\%$  of the GOODS-S field to depths  $\sim 1 \text{ mag}$  shallower than the deep survey (the wide survey also covers three additional fields not used in this study; see

**Table 1**  
Summary of Data—Limiting Magnitudes

Field	Area (arcmin <sup>2</sup> )	$B_{435}$ (mag)	$V_{606}$ (mag)	$i_{775}$ (mag)	$I_{814}$ (mag)	$z_{850}$ (mag)	$Y_{098/105}$ (mag)	$J_{125}$ (mag)	$JH_{140}$ (mag)	$H_{160}$ (mag)
GOODS-S Deep	61.6	28.2	28.6	27.9	28.1	27.8	28.2	28.1	...	27.9
GOODS-S ERS	41.4	28.2	28.5	27.9	27.9	27.6	27.6	28.0	...	27.8
GOODS-S Wide	35.6	28.2	28.7	28.1	27.9	27.9	27.3	27.6	...	27.4
GOODS-N Deep	67.6	28.1	28.3	27.9	...	27.7	28.1	28.3	...	28.1
GOODS-N Wide	71.7	28.1	28.4	27.8	...	27.6	27.3	27.4	...	27.4
HUDF Main	5.1	29.5	30.0	29.7	...	29.1	29.9	29.6	29.6	29.7
HUDF PAR1	4.7	...	29.0	28.8	...	28.5	28.9	29.0	...	28.8
HUDF PAR2	4.8	...	29.0	28.7	...	28.3	28.9	29.2	...	28.9
MACS0416 PAR	4.4	28.8	28.9	...	29.2	...	29.2	29.0	29.0	29.0
Abell 2744 PAR	4.3	29.0	29.1	...	29.2	...	29.1	28.8	28.8	28.9
Zero-points	...	25.68	26.51	25.67	25.95	24.87	26.27	26.23	26.45	25.95

**Note.** The magnitudes quoted are  $5\sigma$  limits measured in  $0''.4$ -diameter apertures on non-PSF-matched images.

Section 6.4.1 and Figure 16). We use ACS imaging from the original GOODS survey in the  $B_{435}$ ,  $V_{606}$ ,  $i_{775}$ , and  $z_{850}$  bands. The most recent ACS mosaics in these fields were used; in GOODS-S this includes all ACS imaging in that field prior to the ACS repair on Servicing Mission 4 in 2009, and the GOODS-N field includes all ACS imaging from the GOODS survey (CANDELS internal team release versions 3 and 2, respectively). The CANDELS imaging in both the deep and wide regions of the GOODS fields includes the  $Y_{105}$ ,  $J_{125}$ , and  $H_{160}$  bands. We add to our GOODS-S data set imaging over the northern  $\sim 25\%$  of the GOODS-S field from the WFC3 Science Oversight Committee’s Early Release Science (ERS) program (PI O’Connell; Windhorst et al. 2011), which also includes  $J_{125}$  and  $H_{160}$  imaging, as well as the F098M (hereafter  $Y_{098}$ ) band. Unless otherwise noted, hereafter we will refer to  $Y_{098}$  and  $Y_{105}$  together as the  $Y$ -band (both filters probe observed  $1\ \mu\text{m}$  light, but the  $Y_{098}$  filter is narrower and thus has a higher spectral resolution).

Finally, we complete our data set with the recently obtained deep *HST* observations near the galaxy clusters Abell 2744 and MACS J0416.1-2403 (hereafter MACS0416) from the HFF program (PI Lotz). For this study, we use only the parallel (unlensed) fields. Both fields have been observed in the  $B_{435}$ ,  $V_{606}$ ,  $I_{814}$ ,  $Y_{105}$ ,  $J_{125}$ ,  $JH_{140}$ , and  $H_{160}$  bands. We use these data to complement our candidate galaxy samples at  $z = 5, 6, 7$ , and 8 (excluding  $z = 4$  due to the reduced number of optical bands).

In parallel to the primary WFC3 observations, CANDELS obtained extremely deep imaging in the F814W band (hereafter  $I_{814}$ ) in both of the GOODS fields. As these data were obtained recently, they suffer from poor charge transfer efficiency. Although algorithms have been devised to correct for this (Anderson & Bedin 2010), we do not include the CANDELS  $I_{814}$  photometry in the initial photometric redshift fitting (although we do explore its inclusion in Section 3.6), because the CANDELS fields have imaging in both the  $i_{775}$  and  $z_{850}$  bands. However, we did use these very deep data during our visual inspection step, which was highly useful at  $z = 8$ , where true  $z = 8$  galaxies should be completely undetected in the  $I_{814}$ -band. In the HFF parallel fields, where the  $I_{814}$  band is the only imaging covering the red end of the optical, we used these data in the full analysis.

The description of the CANDELS *HST* imaging reduction is available from Koekemoer et al. (2011). These reduction steps

were also followed for the ERS, HUDF (Koekemoer et al. 2013), and HFF data we use here. We use imaging mosaics with  $0''.06$  pixels, and make use of their associated weight and rms maps. The combined imaging data set covers an area of  $301.2\ \text{arcmin}^2$ , with  $5\sigma$  limiting magnitudes in the  $H_{160}$  band ranging from 27.4 to 29.7 mag (measured in  $0''.4$  diameter apertures). These data sets are summarized in Table 1.

## 2.2. Point-spread Function (PSF) Matching

The *HST* imaging used here spans more than a factor of three in wavelength, thus the differences in PSF FWHM across that range are significant. For example, the PSF in the GOODS-S Deep field has a FWHM =  $0''.193$  in the  $H_{160}$ -band, but only  $0''.119$  in the  $B_{435}$ -band. A point source will thus have more of its flux contained within a  $0''.4$  aperture in the  $B_{435}$ -band compared with the  $H_{160}$ -band. Because the selection of distant galaxies relies very heavily on accurate colors, and we are using apertures of fixed sizes (determined by the detection image, see Section 2.3) to measure photometry in all bands, the changing PSF needs to be addressed.

We corrected for this by matching the PSF of the *HST* imaging to the  $H_{160}$ -band image (which has the largest PSF FWHM) in each field by using the IDL *deconv\_tool* Lucy–Richardson deconvolution routine in the same way as Finkelstein et al. (2010, 2012b). This routine requires the PSF for a given band, as well as a reference PSF (in this case, the  $H_{160}$ -band), and it generates a kernel. The PSFs were generated by stacking stars in each field in each band, where the stars were selected via identifying the stellar locus in a half-light radius versus magnitude plane. Each star was then visually inspected to ensure that there were no bright near-neighbors, and then the stars were stacked, subsampling by a factor of 10 to ensure an accurate centroiding of each star (i.e., to avoid smearing the PSF during the stacking). Using these PSFs, the deconvolution routine performed an iterative process, relying on the user to determine the number of iterations. We did this by making a guess as to the correct number of iterations, and then changing this number until the stars in the PSF-matched images in a given band had curves of growth that matched the  $H_{160}$ -band curves of growth to within 1% at a

radius of  $0''.4$ . The images were then convolved with the final kernel to generate PSF-matched images.

### 2.3. Photometry

Photometry was measured on the PSF-matched data set with a modified version of the Source Extractor software (v2.8.6, Bertin & Arnouts 1996). Our modified version adds a buffer between the source and the local background cell, and removes spurious sources associated with the distant wings of bright objects. Catalogs were generated independently in each of our 10 subfields, using Source Extractor in two-image mode, where the same detection image was used to measure photometry from all available *HST* filters. For most of our fields, we used a weighted sum of the F125W and F160W images as the detection image to increase our sensitivity to faint objects. In the HUDF main field and the MACS0416 and A2744 HFF parallel fields, we supplemented this catalog with catalogs using 10 additional detection images, derived by stacking all possible combinations of adjacent WFC3 filters. In these three fields, a combined catalog was made up of all unique sources in the catalogs, using a  $0''.2$  matching radius. This allowed very blue sources to be included that may have been too faint in the  $H_{160}$  image to be selected in the original F125W+F160W-selected catalog. This procedure was replicated in our completeness simulations (Section 4). To derive accurate flux uncertainties, Source Extractor relies on both an accurate rms map and a realistic estimate of the effective gain. The provided rms map has been shown to produce accurate uncertainties, and it has been corrected for pixel-to-pixel correlations that occur as a result of the drizzling process (see Guo et al. 2013), which are typically on the order of 10%–15% of the total rms. The effective gains were computed for each band separately, as the the instrument gain (1 for ACS, 2.5 for WFC3/IR) multiplied by the total exposure time for a given image. We have previously verified that the uncertainties measured in this manner on *HST* imaging are accurate (Finkelstein et al. 2012b). The zero-points to convert the observed fluxes into AB magnitudes are given in Table 1, and are appropriate for the dates when these data were taken.

Following our previous work (Finkelstein et al. 2010, 2012a, 2012b, 2013), colors were measured in small Kron apertures with the Source Extractor Kron aperture parameter PHOT\_AUTOPARAMS set to values of 1.2 and 1.7. Finkelstein et al. (2012b) found that these apertures result in more reliable colors for faint galaxies when compared to isophotal or small circular apertures. An aperture correction to the total flux was derived in the H-band and was computed as the ratio between this small Kron aperture flux, and the default Source Extractor MAG\_AUTO flux, which is computed with PHOT\_AUTOPARAMS = 2.5, 3.5. These aperture corrections were then applied to the fluxes in all filters. To see if our aperture corrections accurately recovered the total flux, we examined our completeness simulations (discussed in Section 4) and found that after applying this aperture correction, recovered fluxes were typically 5% fainter in each band than their input fluxes (with the exception of the HUDF main field, where the measured correction was 2%). We thus increased the flux in each band by the appropriate factor to derive our best estimate of the total flux.

The Source Extractor catalogs from each band were combined into a master catalog for each field. At this step, the observed fluxes were corrected for Galactic extinction using

the color excess  $E(B - V)$  from Schlafly & Finkbeiner (2011) appropriate for a given field,<sup>16</sup> and using the Cardelli et al. (1989) Milky Way reddening curve to derive the corrections based on each filter’s central wavelength. We used a mask image to remove objects in regions of bad data, where the mask was generated using a threshold value from the weight map. This mask primarily trims off the noisier edges of the imaging, but it also excludes the “death star” region on the WFC3 array where the number of dithers was low (i.e., in the CANDELS Wide regions). The areas quoted in Table 1 are those of the good regions in these masks. Objects were also removed from the catalog if they had a negative aperture correction, which was the case for a very small number of sources, primarily restricted to areas near very bright sources where the flux in the larger aperture was unreliable. The remaining objects comprised our final catalog in each field.

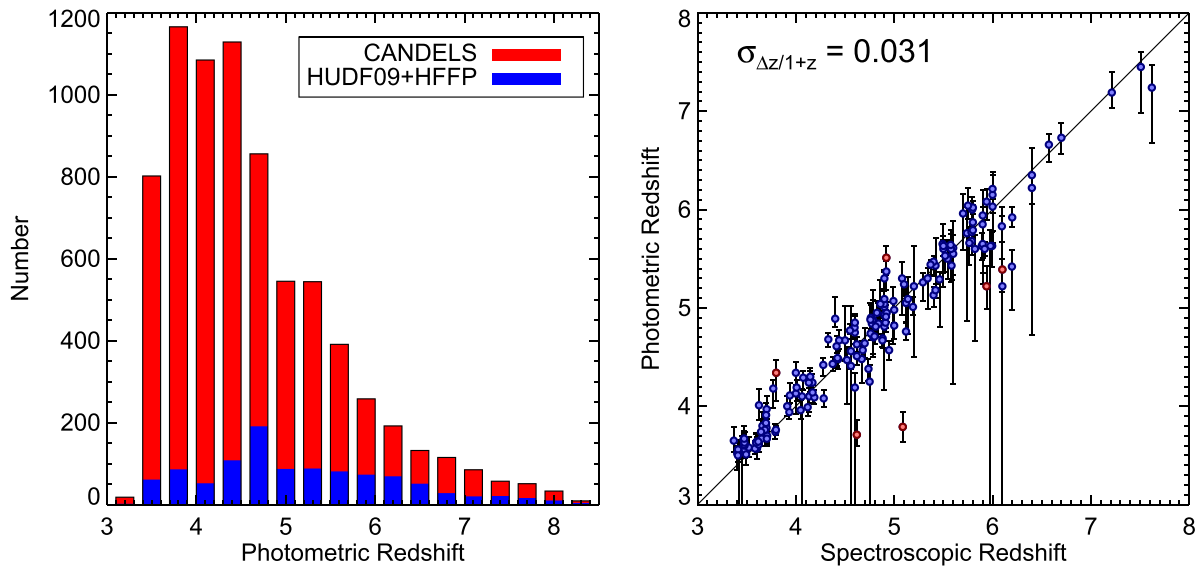
## 3. SAMPLE SELECTION

### 3.1. Photometric Redshifts

We selected our candidate high-redshift galaxy sample via a photometric redshift fitting technique. This has the advantage in that it uses all of the available photometry simultaneously, rather than the multi-step Lyman break galaxy (LBG) method, which selects galaxies using two colors and then subsequently imposes a set of optical non-detection criteria (e.g., Bouwens et al. 2007, 2011b). Another advantage of photometric redshifts is that one obtains a redshift probability distribution function (PDF), which not only allows one to have a better estimate of the redshift uncertainty ( $\sigma_z$  typically  $\sim 0.2$ – $0.3$  versus  $0.5$  for the LBG technique), but can also be used as a tool in the construction of the sample itself. A potential disadvantage of photometric redshift techniques is that the results are based on a set of assumed template spectra; if these templates do not encompass properties similar to the galaxies being studying, systematic offsets may occur (although we also note that similar templates are used to construct LBG color selection criteria). That being said, initial work comparing the differences between galaxy samples selected via LBG and photometric redshift techniques found that the resulting sample properties are fairly similar (e.g., McLure et al. 2013; Schenker et al. 2013).

Photometric redshifts for all sources in the catalogs for each field were measured using the EAZY software (Brammer et al. 2008). The input catalog used all available *HST* photometry, with the exception of the F814W imaging in the CANDELS fields, which was used solely for visual inspection (see Section 2.1). We used an updated set of templates provided with EAZY, based on the PÉGASE stellar population synthesis models (Fioc & Rocca-Volmerange 1997), which now include an increased contribution from emission lines, as recent evidence points to strong rest-frame optical emission lines being ubiquitous among star-forming galaxy populations at high redshift (e.g., Atek et al. 2011; Finkelstein et al. 2011, 2013; van der Wel et al. 2011; Stark et al. 2013; Smit et al. 2014). EAZY assumes the IGM prescription of Madau (1995), and does have the option to include magnitude priors when fitting photometric redshifts, using the luminosity functions as a prior for whether a galaxy at a given apparent magnitude resides at a given redshift. As we show later, there is still non-

<sup>16</sup> <http://ned.ipac.caltech.edu/forms/calculator.html>



**Figure 2.** (Left) the distribution of photometric redshifts in our candidate galaxy sample. The red shading denotes candidates discovered in the CANDELS GOODS fields (including the ERS), while the blue shading denotes those in the combined five deep fields from the HUDF09 and HFF parallel programs. The shallower yet much wider CANDELS imaging dominates the numbers in every redshift bin, by a factor of  $\sim 10$  at  $z = 4-5$ , and  $\sim 2$  at  $z = 7-8$ , although the deep fields are necessary for constraints on the faint end of the luminosity function. (Right) a comparison between the spectroscopic redshift and our best-fit photometric redshifts for the 171 galaxies in our sample with spectroscopic redshifts in the literature. The red circles denote galaxies with  $|z_{\text{spec}} - z_{\text{phot}}| > 1$  at  $\geq 3\sigma$  significance. There are only six such galaxies, and all have spectroscopic redshifts at  $z \gtrsim 4$ .

negligible uncertainty at the bright end of the luminosity function, therefore we did not include these magnitude priors during our photometric redshift fitting process.

### 3.2. Selection Criteria

We selected candidate galaxy samples in five redshift bins centered at  $z \sim 4, 5, 6, 7$ , and  $8$  with  $\Delta z = 1$ , using criteria similar to our previous work (Finkelstein et al. 2012b, 2013). The cosmic time elapsed between our last two bins at  $z \approx 7$  and  $z \approx 8$  is  $\sim 125$  Myr. This time is much longer than the dynamical time of the systems we study, and thus leaves significant time for evolution. However, this will not always be the case as studies of galaxy evolution move toward higher redshift (e.g.,  $\Delta t_{z=13 \rightarrow 12} = 40$  Myr), thus future studies with the *James Webb Space Telescope* (*JWST*) will need to pay careful attention to the choice of sample redshifts when studying galaxy evolution.

Rather than relying solely on the best-fit redshift value, we used the full redshift probability distribution curves  $P(z)$  calculated by EAZY (where  $P(z) \propto \exp(-\chi^2)$ , normalized to unity). Our selection criteria are:

1. A  $\geq 3.5$  significance detection in *both* the  $J_{125}$  and  $H_{160}$  bands. A requirement of a significant detection in two bands removes nearly all spurious sources, since the chances of a noise peak occurring in two images at the same position are very small (Section 3.8.1). This requirement also limits our analysis to galaxies with  $z < 8.5$ , because the Lyman break shifts into the  $J_{125}$  band at  $z = 8.1$ .
2. The integral of the redshift PDF under the primary redshift peak must comprise at least 70% of the total integral. This ensures that no more than 30% of the integrated redshift PDF can be in a secondary redshift solution.

3. The integral under the redshift PDF in the redshift corresponding to a given sample (i.e., 6.5–7.5 for the  $z = 7$  sample) must be at least 25%, which ensures that the redshift PDF is not too broad.
4. The area under the curve in the redshift range of interest must be higher than the area in any other redshift range (i.e., for a galaxy in the  $z = 7$  sample, the integral of  $P(6.5 < z < 7.5)$  must be higher than the integral in any other redshift bin). This criterion ensures that a given galaxy cannot be included in more than one redshift sample.
5. At least 50% of the redshift PDF must be above  $z_{\text{sample}} - 1$  (i.e.,  $\int P(z > 6) > 0.5$  for  $z_{\text{sample}} = 7$ ), and the best-fit redshift must be above  $z_{\text{sample}} - 2$ .
6. The minimum of  $\chi^2$  from the fit must be less than or equal to 60. This criterion ensures that EAZY provides a reasonable fit, although in practice it does not reject many sources.
7. Magnitude in the  $H_{160}$  band must be  $\geq 22$ . This effectively cleans many stars from our sample, but the limit is still more than two magnitudes brighter than our brightest  $z \geq 6$  galaxy candidate. At  $z = 4$ , we do have a few sources close to this limit, but only two sources are brighter than  $H = 22.4$ . This fact—coupled with the observation that the very few sources at  $H < 22$  that satisfy our  $z = 4$  selection criteria are either obvious stars or diffraction spikes—implies that this criterion should not significantly affect our luminosity function results.

Of these criteria, items #1 and #2 are by far the most constraining, as most galaxies that meet these criteria with  $z_{\text{best}} > 3.5$  make it into our sample. Items #3 and #4 are responsible for putting a candidate galaxy in a given redshift sample. While some of these cuts are arbitrary, they will be corrected for when we apply the same criteria to our completeness simulations in Section 4. In Figure 2, we compare the photometric redshifts for 171 galaxies in our

sample to available spectroscopic redshifts in the literature.<sup>17</sup> The agreement is excellent, with  $\sigma_{\Delta z/(1+z)} = 0.031$  (derived by taking an iterative  $3\sigma$ -clipped standard deviation), although the number of confirmed redshifts at  $z > 6.5$  is small (only five galaxies). The number of outliers is also small, with only six out of 171 galaxies (3.5%) having a photometric redshift differing from the spectroscopic redshift by  $\Delta z > 1$  at  $\geq 3\sigma$  significance. All six galaxies have  $z_{\text{spec}} \gtrsim 4$ , thus no galaxies in our sample have a catastrophically lower spectroscopic redshift. In comparison, when defining outliers in the same way, we find that the published CANDELS team photometric redshift catalog has 13 outliers out of 174 total spectroscopic redshifts, for a somewhat higher outlier fraction of 7.5% (Dahlen et al. 2013). Although the fraction of galaxies with confirmed redshifts is relatively small, the available spectroscopy confirms that our selection methods yield an accurate high-redshift sample. In the remainder of this paper, we will therefore refer to our candidate galaxies solely as galaxies, with the caveat that spectroscopic follow up of a much larger sample, particularly at  $z > 6$ , is warranted.

### 3.3. Visual Inspection

The candidate selection process is automated, so for a truly robust galaxy sample we required a visual inspection of each of our  $\sim 7500$  candidate high-redshift galaxies. During the visual inspection, we examined the following features:

1. Is the source a real galaxy? Objects were inspected to ensure that they were not an artifact, such as part of a diffraction spike (which frequently appear in different places in the ACS and WFC3 imaging due to different roll angles during the respective observations), oversplit regions of bright galaxies, or noise near the edge of the images.
2. Is the aperture drawn correctly? While the small Kron apertures yield the most reliable colors, they are also susceptible to “stretching” (i.e., becoming highly elongated) in regions of high noise or near very bright objects. For each source, we compared the ratio of the flux between the Kron aperture and a  $0''.4$ -diameter circular aperture to that same quantity for objects of a similar magnitude from the full photometry catalog. If an object had a value  $\gtrsim 30\%$  higher than similarly bright sources in the full photometry catalog *and* the aperture looked affected by noise/bright sources, we adjusted the photometry of the object in question accordingly, using the  $0''.4$ -to-total correction of similarly bright objects in the catalog. In practice, these issues affected  $< 10\%$  of the galaxies in our high-redshift sample.
3. Is there significant optical flux that did not get measured correctly? Primarily due to the issues with the inaccurate apertures discussed previously, a very small number of sources appeared to have optical flux when visually inspected that was not measured to be significant in our catalog (i.e., in the case of a too-large aperture, the flux is concentrated in a small number of pixels, whereas the

flux error comes from the full aperture, so the signal-to-noise is low). In these cases, objects were removed from our sample. This step is somewhat qualitative, as there are cases of objects where the aperture appears correct, yet there is still a  $\sim 1$ - $2\sigma$  detection in a single optical band. In the majority of these cases, we left these objects in the sample because we are confident in our photometric redshift analysis. During this step, we also examined  $I_{814}$  photometry for each source in the CANDELS fields; this primarily benefits the selection of  $z = 8$  galaxies, which should not be visible at this wavelength. Three  $z = 8$  candidates with observable  $I_{814}$  flux were removed from our sample.

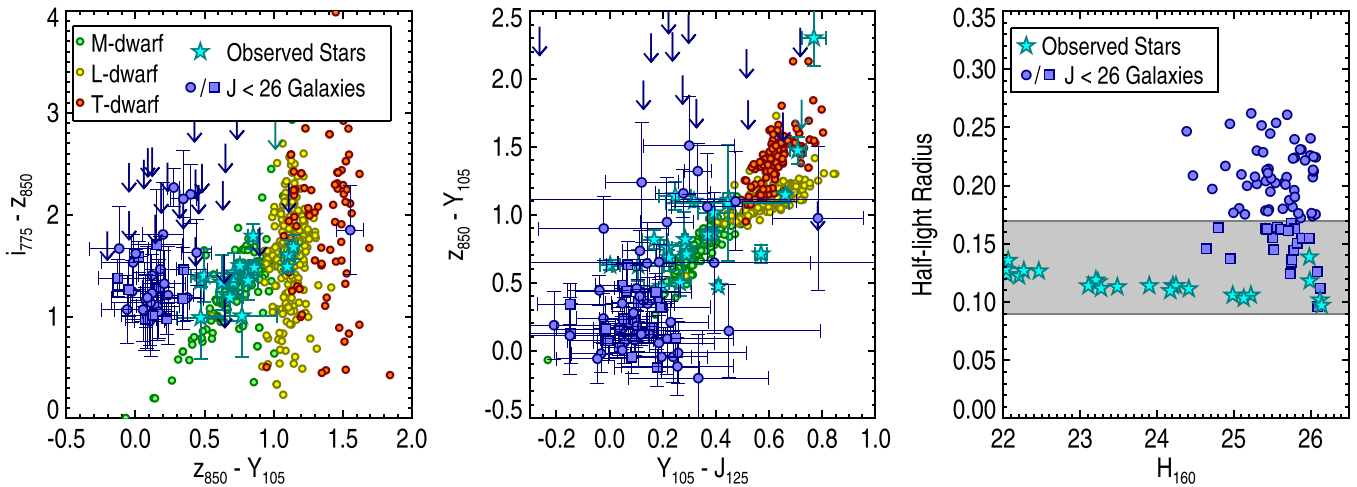
### 3.4. Stellar Contamination

The most crucial step in our visual inspection is the classification and removal of stellar sources, because stellar contamination would dominate the bright end of the luminosity function if these contaminants were not considered. In particular, M-type stars and L- and T-brown type dwarf stars can have similar colors (including optical non-detections) as our high-redshift galaxies of interest, especially at  $z \geq 6$ . While some studies use dwarf star colors during their selection (e.g., Bowler et al. 2012, 2014; McLure et al. 2013; Bouwens et al. 2015), many primarily use the Source Extractor “stellarity” parameter to diagnose whether a compact object is a star or a galaxy (e.g., Bouwens et al. 2011b, 2015). However, the stellarity parameter loses its ability to discern between a point source and a resolved source for faint objects. To test this further, we examined the stellarity of sources in the CANDELS GOODS catalogs. At very bright magnitudes ( $J_{125} < 24$ ), there is a clear separation between stars and galaxies, with objects either having a stellarity near unity (i.e., stars) or a stellarity near zero (i.e., galaxies). However, this separation becomes less clear at  $J_{125} > 25$ , where the stellar and galaxy sequences begin to blend together. Therefore, stellarity can be an unreliable star-galaxy separator at  $J_{125} > 25$ , which is similar to the brightness of our brightest  $z \geq 7$  galaxies.

While the GOODS fields cover relatively small regions on the sky, the potential number of brown dwarf contaminants, even at  $J_{125} > 25$ , is significant. The Galactic structure model of R. Ryan & N. Reid (2015, in preparation) predicts the surface density of brown dwarfs in our covered fields. In the GOODS-S region, using the area covered by the CANDELS, ERS, and HUDF09 observations, we would expect  $\sim 6$  stars of spectral type M6–T9 with  $J_{125}$ -band magnitudes between 25 and 27. The surface density of M6–T9 stars in GOODS-N is similar, with an expected number of stars in the field of  $\sim 5$ . Thus, the expected number of  $25 < J_{125} < 27$  stars of spectral type M6–T9 in our whole surveyed region is  $\sim 11$ . While this number is small, the number of brown dwarfs is expected to fall off toward fainter magnitudes, thus the majority of these likely have  $J_{125}$  close to 25. This magnitude is similar to those of the brightest galaxies in our sample, which dominate the shape of the bright end of the galaxy luminosity function. As stellarity is an unreliable method of identifying these sources, we must find an alternative method.

Although brown dwarfs can have similar colors to  $z > 6$  galaxies, and can be included in the initial sample, they fall on well-defined color sequences, and can thus be distinguished

<sup>17</sup> The spectroscopic redshifts come from a compilation made by N. Hathi (2015, private communication), which include data from the following studies: Szokoly et al. (2004), Grazian et al. (2006), Barger et al. (2008), Hathi et al. (2008), Vanzella et al. (2008, 2009), Rhoads et al. (2009), Wuyts et al. (2009), Balestra et al. (2010), Ono et al. (2012), Finkelstein et al. (2013), Kurk et al. (2013), and Rhoads et al. (2013).



**Figure 3.** Left and center: color–color plots. Blue circles and squares denote objects accepted as  $J < 26$ ,  $z \gtrsim 6$  galaxies, with squares indicating the ones with half-light radii  $< 0''.17$ . Arrows represent  $1\sigma$  limits. Cyan stars denote candidates originally selected as galaxies but reclassified as stars based on their sizes and colors. Small circles denote known stars from the the SpeX Prism Spectral Libraries from the 3 m NASA Infrared Telescope Facility, with spectral types as indicated in the legend. Right: half-light radius vs. magnitude for  $J < 26$  candidates. Symbols are the same as in the other panels. No compact galaxies have colors similar to known stars in both color–color plots. Similar plots were used to exclude stellar contaminants at  $z = 4$  and 5.

from true galaxies. Figure 3 shows the two color–color plots,<sup>18</sup> which we used in tandem with the size information—examining not only stellarity, but also the FWHM and half-light radius as measured by Source Extractor—to identify stars lurking our sample (similar plots were used at  $z = 4$  and 5). If a galaxy appeared un-resolved (defined as having a stellarity  $> 0.8$ , or a half-light radius and/or FWHM similar to that of stars in the field), then we examined that object in the color–color plots as shown in Figure 3. If the object also had colors similar to a dwarf star, we removed it from our sample. Over all of our fields, a total of 23 objects were flagged as stars (many with  $J < 25$ ) in our  $z \gtrsim 6$  samples: 18 from our initial  $z \sim 6$  galaxy sample and five from our initial  $z \sim 7$  galaxy sample. These objects were removed from our sample. One of the stars removed from our  $z \sim 7$  sample was previously flagged as a probable T-dwarf by Castellano et al. (2010). We examined a subset of eight of these stars, which were detected in the FourStar Galaxy Evolution (zFourGE) medium band imaging survey of a portion of GOODS-S, and found that all eight have  $z - J_1$  and  $J_1 - J_3$  colors consistent with brown dwarf stars (where  $J_1$  and  $J_3$  refer to two of the three medium bands that comprise the  $J$  band; Tilvi et al. 2013). Of these, six stars have  $J_{125} > 25$ , meaning that our high-redshift galaxy selection criteria also originally selected about half of the expected number of faint brown dwarfs in this field. Four of these six stars have Source Extractor stellarity measurements  $< 0.8$ , thus a stellarity-only rejection method would have failed to remove them. We conclude that our visual inspection step efficiently removed stellar contaminants from our sample, but we emphasize that the color examination portion was crucial to exclude the faintest stars from our sample.

### 3.5. Active Galactic Nuclei

We screened for the presence of bright AGN in our sample by searching for counterparts in *Chandra X-ray Observatory* point source catalogs. In the GOODS-S field, we used the 4 Msec *Chandra* Deep Field—South (CDF-S) catalog of Xue

et al. (2011); in GOODS-N, we used the 2Msec *Chandra* Deep Field—North catalog of Alexander et al. (2003). These catalogs have average positional accuracies of  $0''.42$  and  $0''.3$ , respectively. To be conservative, we searched for matches in each catalog out to a radius of  $1''$ . We then visually inspected each of the 34 galaxies in our sample with a match. Seven objects, all with *Chandra* catalog separations  $> 0''.6$ , had nearby *HST* counterparts with positions that were consistent with the *Chandra* catalog. These seven sources remained in our sample, because these interlopers were likely providing the X-ray emission, and none had spectroscopic redshifts in the CDF-S catalog. The remaining 27 sources, all with *Chandra* catalog separations  $\leq 0''.6$  from our candidate galaxies had *HST* counterparts with positions consistent with the X-ray emission coming from the galaxies in our sample. Secure spectroscopic redshifts were available for four of these 27 galaxies in the CDF-S catalog:  $z = 3.06, 3.66, 3.70,$  and  $4.76$ . We removed these 27 galaxies (25 from our  $z = 4$  sample, and two from our  $z = 5$  sample) from our galaxy sample. This removal is conservative, because, although the X-ray detections imply the presence of an AGN, they do not prove that the AGN dominates the UV luminosity.

### 3.6. Photometric Redshifts with Spitzer/IRAC Photometry

As we will discuss later, one of the main results of this work is an apparent constant value of  $M_{UV}^*$  at  $z > 5$ , which is brighter than many previous works. It is thus imperative that we have high confidence that our bright galaxies are all actually at high redshift, and not lower-redshift contaminants. To provide a further check on our bright sources, we re-examined the photometric redshifts of our bright galaxies with the addition of *Spitzer Space Telescope* Infrared Array Camera (IRAC; Fazio et al. 2004) imaging over our fields. This imaging probes the rest-frame optical at these wavelengths, and thus provides significant constraining power because the most likely contaminants are red, lower-redshift galaxies, which would have very different fluxes in the mid-infrared than true high-redshift galaxies. We examined sources with  $M_{1500} < -21$ , which is approximately the value of  $M_{UV}^*$  at these redshifts, and

<sup>18</sup> This research benefitted from the SpeX Prism Spectral Libraries, maintained by Adam Burgasser at <http://pono.ucsd.edu/~adam/browndwarfs/spexprism>.

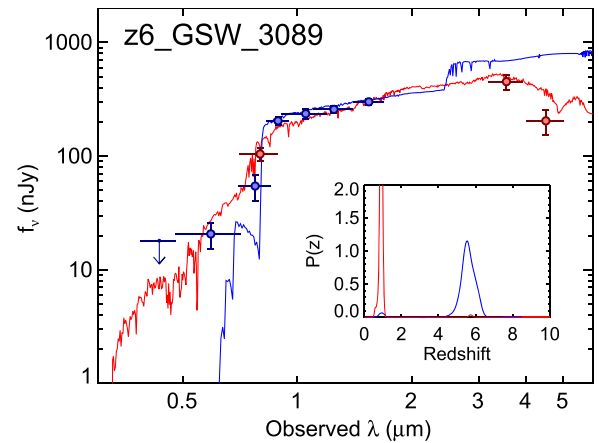


provides samples of 164, 85, 29, 18, and 3 bright galaxies at  $z = 4, 5, 6, 7,$  and  $8,$  respectively.

During the cryogenic mission, the GOODS fields were observed by the GOODS team (M. Dickinson et al. 2015, in preparation) at  $3.6, 4.5, 5.8,$  and  $8.0 \mu\text{m}.$  Later, during Cycle 6 of the warm mission, broader regions encompassing the GOODS footprints were covered by the *Spitzer* Extended Deep Survey (SEDS Ashby et al. 2013) to  $3\sigma$  depths of 26 AB mag at both  $3.6$  and  $4.5 \mu\text{m}.$  A somewhat narrower subset of both fields was subsequently covered by *Spitzer*-CANDELS (S-CANDELS Ashby et al. 2015) to even fainter levels; reaching  $\sim 0.5$  mag deeper than SEDS in both of the warm IRAC bandpasses. The HUDF09 fields were observed by *Spitzer* program 70145 (the IRAC Ultra-Deep Field; Labbé et al. 2013), reaching 120, 50, and 100 hr in the HUDF Main, PAR1, and PAR2 fields, respectively. Finally, program 70204 (PI Fazio) observed a region in the ERS field to 100 hr depth. The present work is based on mosaics constructed by coadding all the above data, following the procedures described by Ashby et al. (2013). The combined data have a depth of  $\gtrsim 50$  hr over both CANDELS GOODS fields and  $>100$  hr over the HUDF main field.

Because the IRAC PSF is much broader than that of *HST*, our galaxies may be blended with other nearby sources. We measure *Spitzer*/IRAC  $3.6$  and  $4.5 \mu\text{m}$  photometry by performing PSF-matched photometry on the combined IRAC data, which reach at least 26.5 mag ( $3\sigma$ ) at  $3.6$  and  $4.5 \mu\text{m}$  (Ashby et al. 2015). We utilized the TPHOT software (Merlin et al. 2015), an updated version of TFIT (Laidler et al. 2007), to model low-resolution images (IRAC images) by convolving *HST* imaging with empirically derived IRAC PSFs and simultaneously fitting all IRAC sources. Specifically, we used the light profiles and isophotes in the detection ( $J+H$ ) image obtained by Source Extractor, and convolved them with a transfer kernel to generate model images for the low-resolution data. These models were then fit to the real low-resolution images, dilating the segmentation maps of the model images to account for missing flux on the edges of galaxies (Galamez et al. 2013). The fluxes of sources are determined by the model that best represents the real data. As the PSF FWHM of the high-resolution image (H-band) is negligible ( $\sim 0''.19$ ) when compared to those of the low-resolution IRAC images ( $\sim 1''.7$ ), we use the IRAC PSFs as transfer kernels. We derive empirical IRAC PSFs by stacking isolated and moderately bright stars in each field. Because our own WFC3 catalog was used as the input for TPHOT, all of our galaxies have IRAC measurements in the TPHOT catalogs. We visually inspected the positions of each of our high-redshift galaxy candidates in the IRAC images to ensure no significant contamination from the residuals of nearby bright galaxies. If an object was on or near a strong residual, we ignored the IRAC photometry in the subsequent analysis. This was the case for 18 of the 164 galaxies at  $z = 4,$  23 of the 85 galaxies at  $z = 5,$  3 of the 29 galaxies at  $z = 6,$  6 of the 18 galaxies at  $z = 7,$  and 1 of the 3 galaxies at  $z = 8.$  With the contaminated fluxes removed, we found that all remaining  $M_{1500} < -21$  galaxies at  $z = 4-8$  had  $3.6 \mu\text{m}$  detections of at least  $3\sigma$  significance, with a magnitude range at  $z \geq 6$  of  $22.7 \leq m_{3.6} \leq 25.8.$  The full description of our TPHOT IRAC photometry catalog will be presented by Song et al. (2015).

We reran EAZY for this subsample of bright galaxies, including the *Spitzer*/IRAC fluxes, as well as photometry from the ACS F814W filter, which was not included in the original



**Figure 4.** SED of the only galaxy in our 50-object sample of bright ( $M_{1500} \leq -21$ )  $z \gtrsim 6$  galaxies that had a photometric redshift that preferred a low-redshift solution after the inclusion of IRAC and F814W photometry. The blue curve shows the original high-redshift best-fitting stellar population model and photometric redshift probability distribution function, while the red curve shows the results including IRAC and F814W. This galaxy was removed from our sample, because the IRAC photometry is consistent with the stellar emission peak at  $z \sim 1.$  The inferred contamination rate of 2% (one out of 50 galaxies) is even lower than our estimates for  $z \gtrsim 6$  in Section 3.8.

photometric redshift calculation (see Section 2.1). We examined these updated photometric redshift results, searching for galaxies in our  $z = 4$  and  $5$  samples with  $z_{\text{new}} < 2.5,$  and in our  $z = 6, 7,$  and  $8$  samples with  $z_{\text{new}} < 4.$  We found 14 out of 164 galaxies in our  $z = 4$  sample, and 14 out of 85 galaxies in our  $z = 5$  sample with  $z_{\text{new}} < 2.5.$  We found 1 galaxy out of 29 at  $z = 6$  that appeared to be better fit with a low-redshift solution of  $z_{\text{new}} = 0.9,$  but no galaxies in our  $z = 7$  or  $8$  samples had preferred low-redshift solutions with the inclusion of IRAC photometry.

Examining these results, out of the 28  $z = 4$  or  $5$  galaxies with preferred low-redshift solutions, 23 had photometry consistent with a true low-redshift galaxy. Four galaxies, however, had photometry that appeared to be consistent with a high-redshift galaxy with a strong emission line ( $H\alpha$  or  $[O III]$ ) in one IRAC band. Systems with lines such as these (i.e.,  $EW_{[O III]} > 500 \text{ \AA}$ ) are rare locally, but appear to be more common at high redshift (e.g., van der Wel et al. 2011; Finkelstein et al. 2013; Smit et al. 2014). Although typical emission line strengths are now included in the EAZY templates, these do not account for extreme emission lines, so it is not surprising that EAZY does not return a high-redshift solution. We elect to keep these four galaxies in our sample, noting that the lack of strong rest-frame optical lines in the EAZY templates does not affect our initial sample selection, which does not make use of the IRAC photometry. A fifth galaxy (z5\_GNW\_13415) has a high-quality published spectroscopic redshift of  $z = 5.45,$  thus we also keep it in our sample. Perhaps unsurprisingly, the low-redshift fit for this galaxy had a very poor quality-of-fit, with  $\chi^2 = 127,$  implying that the EAZY templates are a poor match for this galaxy. The remaining 23 galaxies at  $z = 4$  and  $5$  were removed from our sample. The sole  $z \geq 6$  galaxy with a preferred low-redshift solution with the inclusion of IRAC photometry, z6\_GSW\_3089, is shown in Figure 4. The red *HST* colors imply a much brighter IRAC flux than is seen. A solution at  $z = 0.93$  yields a better fit, because the peak of stellar emission at that redshift better matches the observed IRAC fluxes. We

**Table 2**  
Summary of Final High-redshift Galaxy Samples

Redshift	$N_{\text{all}}$	$N_{M < -21}$	$V_{\text{eff}} (M_{1500} = -22)$ ( $10^5 \text{ Mpc}^3$ )	$V_{\text{eff}} (M_{1500} = -19)$ ( $10^5 \text{ Mpc}^3$ )
4 (3.5–4.5)	4156	150	12.2	4.11
5 (4.5–5.5)	2204	77	8.98	3.36
6 (5.5–6.5)	706	28	7.93	2.50
7 (6.5–7.5)	300	18	6.99	0.30
8 (7.5–8.5)	80	3	5.88	0.16

**Note.** The total number of sources in our final galaxy sample, after all contaminants were removed. The final two columns give the total effective volume at each redshift for two different values of the UV absolute magnitude.

thus removed this galaxy from our sample. This galaxy also has a spectroscopic redshift of  $z = 5.59$  from the observations of Vanzella et al. (2009); however, it has a spectroscopic quality flag of “C,” which indicates that the spectroscopic redshift is unreliable. This combined with the  $\sim 4\sigma$  detection in the  $V_{606}$  band leaves us confident that a low-redshift solution is more likely.

With the removal of these likely contaminants, we retain a total sample of 150, 77, 28, 18, and 3  $M_{1500} < -21$  galaxies at  $z = 4, 5, 6, 7,$  and  $8,$  respectively. The fraction of contaminants at  $z \geq 6$  (1 out of 50  $z = 6, 7,$  and  $8$  galaxies, or 2%) is consistent with (albeit somewhat less than) the expected low value calculated in Section 3.8.

Our final galaxy sample is summarized in Table 2, and a catalog of all galaxies in our sample is provided in Table 3. In Figure 1 we show the absolute UV magnitude distribution of our final samples, highlighting that we cover a dynamic range of five magnitudes. In particular, the CANDELS data are crucial, because galaxies from these data dominate the total number of galaxies in our sample, and approximately double the luminosity dynamic range that we can probe. This is highlighted in the left panel of Figure 2, which shows that galaxies discovered in the CANDELS GOODS fields dominate the total number at all redshifts in our sample.

### 3.7. Stellar Population Modeling

To derive the rest-frame absolute magnitude at  $1500 \text{ \AA}$  ( $M_{1500}$ ), as well as the UV spectral slope  $\beta$  ( $f_{\lambda} \propto \lambda^{\beta}$  Calzetti et al. 1994), we fit spectral energy distributions (SEDs) from synthetic stellar population models to the observed *HST* photometry of our high-redshift candidate galaxies. The technique used here is similar to our previous works (Finkelstein et al. 2010, 2012a, 2012b, 2013). We used the updated (2007) stellar population synthesis models of Bruzual & Charlot (2003) to generate a grid of spectra, varying the stellar population metallicity, age, and star-formation history (SFH).<sup>19</sup> Metallicities spanned 0.02–1 times  $z_{\odot}$ , and ages spanned 1 Myr to the age of the universe at a given redshift. We allowed several different types of SFHs, including a single burst and continuous, as well as both exponentially decaying

<sup>19</sup> These models may overestimate the contribution of thermally pulsating asymptotic giant branch (TP-AGB) stars. However, these stars typically begin to dominate the emission at population ages  $\gtrsim 1$  Gyr. Additionally, although the TP-AGB contribution may affect the SED in post-starburst galaxies at wavelengths as low as  $0.5 \mu\text{m}$  (Kriek et al. 2010), our longest wavelength filter ( $1.6 \mu\text{m}$ ) at our lowest redshift ( $z = 4$ ) probes only  $0.3 \mu\text{m}$ , and all other filter/redshift combinations probe bluer rest-frame wavelengths. Thus, our choice to use the updated models should have no effect on our results.

and rising (so-called “tau” and “inverted-tau” models). To these spectra, we added dust attenuation using the starburst attenuation curve of Calzetti et al. (2000), with a range of  $0 \leq E(B - V) \leq 0.8$  ( $0 \leq A_V \leq 3.2$  mag). We also included nebular emission lines using the prescription of Salmon et al. (2015)—which uses the line ratios from Inoue (2011), based on the number of ionizing photons from a given model—and assuming the ionizing photon escape fraction is  $\approx$  zero. We then redshifted these models to  $0 < z < 11$  and added IGM attenuation (Madau 1995). These model spectra were integrated through our *HST* filter bandpasses to derive synthetic photometry for comparison with our observations. For each model, we computed the value of  $M_{1500}$  by fitting a  $100 \text{ \AA}$ -wide synthetic top-hat filter to the spectrum centered at rest-frame  $1500 \text{ \AA}$ . Likewise, for each model we measured the value of  $\beta$  by fitting a power law to each model spectrum using the wavelength windows specified by Calzetti et al. (1994), similar to Finkelstein et al. (2012b).

The best-fit model was found via  $\chi^2$  minimization, including an extra systematic error term of 5% of the object flux for each band to account for such items as residual uncertainties in the zeropoint correction and PSF-matching process. The stellar mass was computed as the normalization between the best-fit model (which was normalized to  $1 M_{\odot}$ ) and the observed fluxes, weighted by the signal-to-noise in each band. These best-fit values of  $M_{1500}$  are used in our luminosity function analysis, while  $\beta$  is used to correct for incompleteness in a color-dependent fashion. The uncertainties in the best-fit parameters were derived via Monte Carlo simulations, perturbing the observed flux of each object by a number drawn from a Gaussian distribution with a standard deviation equal to the flux uncertainty in a given filter. For each galaxy,  $10^2$  Monte Carlo simulations were run, providing a distribution of  $10^2$  values for each physical parameter. The 68% confidence range for each parameter was calculated as the range of the central 68% of results from these simulations. The best-fit redshift in these simulations was allowed to vary following the redshift PDF, thus folding in the uncertainty in redshift into the uncertainty in the physical parameters (most notably, the stellar mass and  $M_{1500}$ ; Finkelstein et al. 2012a). During this process, we only allowed the redshift to vary within  $\Delta z = \pm 1$  of the best-fit photometric redshift. This excludes any low-redshift solution from biasing the uncertainties on a given parameter. The amount of the integrated  $P(z)$  at  $z > 3$  excluded via this step was typically  $\leq 10\%$  (at  $z = 6$ ).

### 3.8. Contamination

A key issue in any study of high-redshift galaxies is the risk of sample contamination, either by spurious sources or by lower-redshift interlopers. The gold standard for eliminating contamination is to obtain spectroscopic redshifts. This is clearly unfeasible for all galaxies in our sample (until the next generation of space and ground-based telescopes), but there is significant archival spectroscopic data. As discussed in Section 3.2, we find excellent agreement between available spectroscopic redshifts and our photometric redshifts, with  $\sigma_{z/(1+z)} = 0.031$ . In particular, the four bright<sup>20</sup> galaxies (

<sup>20</sup> Object z7\_MAIN\_2771 has a tentative spectroscopic redshift of  $z_{\text{spec}} = 7.62$ , based on a  $4\sigma$  possible Ly $\alpha$  emission line from Schenker et al. (2014). This object is quite faint, with  $J_{125} = 28.8$ , resulting in a somewhat large 95% photometric redshift confidence range of 6.02–7.74, although consistent with the tentative spectroscopic redshift.

**Table 3**  
Catalog of Candidate Galaxies at  $3.5 \lesssim z \lesssim 8.5$

Catalog ID	IAU Designation	R.A. (J2000)	Decl. (J2000)	$z_{\text{phot}}$	$M_{1500}$ (AB mag)
z4_GSD_27037	HRG14 J033240.8–275003.1	53.169922	–27.834183	3.54 (3.45 to 3.63)	–20.45 (–20.57 to –20.42)
z4_ERS_3675	HRG14 J033235.0–274117.5	53.145882	–27.688189	4.08 (3.79 to 4.28)	–20.39 (–20.47 to –20.15)
z4_GND_29830	HRG14 J123718.1+621309.7	189.325211	62.219368	4.01 (3.85 to 4.17)	–19.68 (–19.81 to –19.54)
z4_GND_30689	HRG14 J123721.4+621259.2	189.339355	62.216450	3.66 (3.59 to 3.75)	–21.09 (–21.16 to –21.03)
z5_GSD_8969	HRG14 J033216.2–274641.6	53.067379	–27.778219	5.00 (4.87 to 5.14)	–20.62 (–20.68 to –20.51)
z5_GND_31173	HRG14 J123731.0+621254.2	189.379272	62.215046	4.85 (4.37 to 5.09)	–19.59 (–19.77 to –19.43)
z5_MAIN_3271	HRG14 J033243.5–274711.4	53.181351	–27.786510	5.50 (4.58 to 5.67)	–16.95 (–17.05 to –16.78)
z5_PAR2_3762	HRG14 J033304.8–275234.7	53.270004	–27.876295	4.47 (3.71 to 4.70)	–18.83 (–18.97 to –18.57)
z6_GND_16819	HRG14 J123718.8+621522.7	189.328232	62.256317	5.55 (5.41 to 5.65)	–21.56 (–21.62 to –21.47)
z6_GNW_16070	HRG14 J123549.0+621224.8	188.954025	62.206898	5.88 (5.64 to 6.02)	–20.83 (–20.88 to –20.64)
z6_MAIN_2916	HRG14 J033244.8–274656.8	53.186806	–27.782433	6.42 (5.79 to 6.76)	–18.39 (–18.55 to –18.19)
z6_MACS0416PAR_145	HRG14 J041632.2–240533.3	64.134117	–24.092587	5.91 (5.08 to 6.23)	–18.49 (–18.64 to –18.16)
z7_GSD_12285	HRG14 J033206.7–274715.8	53.028114	–27.787714	7.30 (6.44 to 7.89)	–19.57 (–19.79 to –19.31)
z7_ERS_6730	HRG14 J033216.0–274159.2	53.066677	–27.699766	6.74 (5.64 to 6.87)	–20.31 (–20.31 to –19.96)
z7_GND_16759	HRG14 J123619.2+621523.2	189.079834	62.256454	6.69 (6.33 to 6.89)	–20.89 (–20.98 to –20.76)
z7_A2744PAR_4276	HRG14 J001357.5–302358.3	3.489512	–30.399530	6.51 (6.26 to 6.78)	–19.37 (–19.47 to –19.22)
z8_GSD_16150	HRG14 J033213.9–274757.7	53.057983	–27.799349	7.91 (6.21 to 8.54)	–20.14 (–20.38 to –19.91)
z8_MAIN_5173	HRG14 J033241.5–274751.0	53.172874	–27.797487	8.11 (6.29 to 8.64)	–17.67 (–17.98 to –17.41)
z8_GND_32082	HRG14 J123727.4+621244.4	189.364258	62.212334	7.64 (7.02 to 8.16)	–20.27 (–20.33 to –20.02)
z8_GND_8052	HRG14 J123704.8+621718.8	189.270020	62.288559	8.10 (7.04 to 8.41)	–20.68 (–20.84 to –20.44)

**Note.** A catalog of our 7446  $z = 4$ –8 galaxy candidates, with their derived properties. We include the IAU designation for continuity with previous and future works, with a designation prefix HRG14 denoting “High Redshift Galaxy 2014.” The values in parentheses represent the 68% confidence range on the derived parameters. Here, we show 20 representative galaxies, four from each redshift bin. A portion is shown here for guidance regarding its form and content.

(This table is available in its entirety in machine-readable form.)

$24.9 < J_{125} < 25.7$ ) with confirmed  $z_{\text{spec}} > 6.5$  have an excellent agreement with spectroscopic redshifts: z7\_GNW\_24443 with  $z_{\text{phot}} = 6.66$  and  $z_{\text{spec}} = 6.573$  (Rhoads et al. 2013); z7\_GSD\_21172 with  $z_{\text{phot}} = 6.73$  and  $z_{\text{spec}} = 6.70$  (Hathi et al. 2008); z7\_GNW\_4703 with  $z_{\text{phot}} = 7.19$  and  $z_{\text{spec}} = 7.213$  (Ono et al. 2012); and z7\_GND\_42912<sup>21</sup> with  $z_{\text{phot}} = 7.45$  and  $z_{\text{spec}} = 7.51$  (Finkelstein et al. 2013). The brightest source in our  $z = 6$  sample, z6\_GSW\_12831 with  $M_{1500} = -22.1$  and  $z_{\text{phot}} = 5.77$ , is confirmed with  $z_{\text{spec}} = 5.79$  (Bunker et al. 2003). This galaxy has a  $3\sigma$  detection in the  $V_{606}$ -band, which could have resulted in its exclusion from a typical LBG color–color selection sample, although the observed flux can be explained by non-ionizing UV photons transmitted through the Ly $\alpha$  forest.

In general, spectroscopic follow up of sources selected on the basis of their Lyman breaks (either color–color selection, or photometric redshift selection) finds a very small contamination rate by low-redshift sources (e.g., Pentericci et al. 2011). However, given the apparent difficulty in detecting Ly $\alpha$  emission at  $z > 6.5$  (e.g., Pentericci et al. 2011; Ono et al. 2012; Finkelstein et al. 2013), the true effect of contamination at these higher redshifts is not empirically known. In this subsection, we will attempt to estimate our contamination fraction by other means.

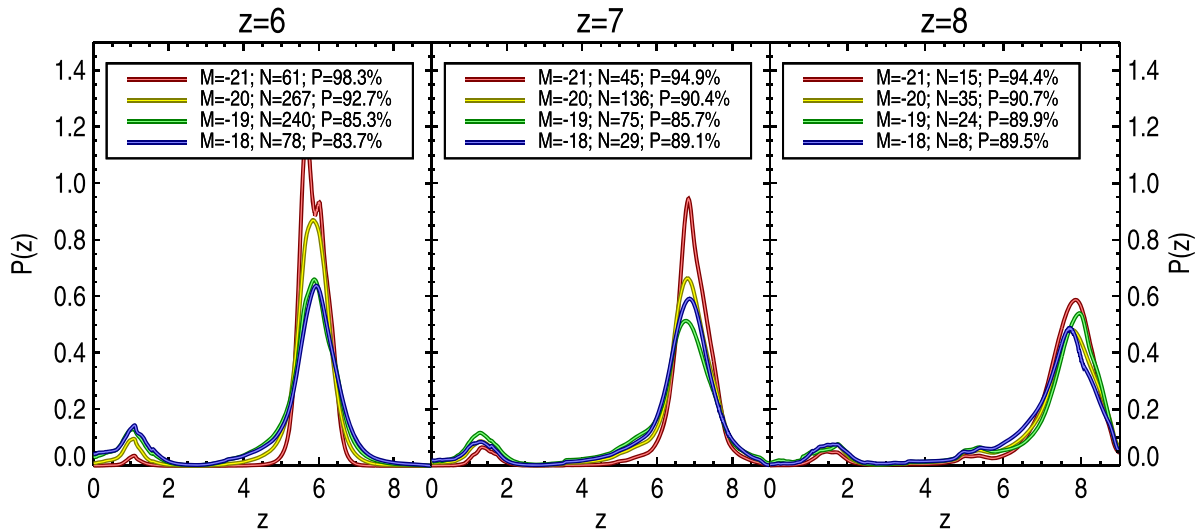
<sup>21</sup> This source was originally called z8\_GND\_5296 in our previous catalog (Finkelstein et al. 2013). Our new catalog uses an updated version of the CANDELS GOODS-N data, thus the catalog numbering is different. Additionally, the slightly updated photometry pushes the photometric redshift of this galaxy slightly below  $z = 7.5$ , placing it in the  $z = 7$  sample.

### 3.8.1. Properties of the Image Noise

Two key components of our selection processes should eliminate contamination by spurious sources in our sample. First, we restricted our sample to galaxies detected at  $\geq 3.5\sigma$  in two imaging bands:  $J_{125}$  and  $H_{160}$ . Formally, requiring a  $3.5\sigma$  detection in a single band should yield only a 0.05% contamination by noise. However, the wings of the noise distribution are highly non-Gaussian. We examined this by measuring the fluxes at  $2 \times 10^5$  random positions in the  $J_{125}$  and  $H_{160}$  images in the GOODS-S Deep field (see Schmidt et al. 2014 for a similar analysis). To avoid biasing from real objects in the image, we only considered negative fluctuations (e.g., Dickinson et al. 2004), where the contamination percentage was computed as the ratio of the number of apertures with a flux  $< -1 \times 3.5\sigma$  to the total number of apertures with negative fluxes. We found that in each of these bands individually, the fraction of positions measuring at  $> 3.5\sigma$  was  $\sim 1.4\%$ , which is much higher than predicted based on an assumption of Gaussian noise. If we instead look at the number of  $3.5\sigma$  fluctuations in both the  $J_{125}$  and  $H_{160}$  images at the same location, we find that this contamination drops to nearly zero, at 0.05%. Thus, we conclude that because we require significant detections in two bands, the contamination in our sample by noise is negligible. Spurious sources other than noise spikes are eliminated by our detailed visual inspection of each source, as described in Section 3.3.

### 3.8.2. Estimates from Stacked Redshift PDFs

Contamination by low-redshift interlopers is a more complicated issue. While extreme emission line galaxies at lower redshift could theoretically be an issue (Atek et al. 2011),



**Figure 5.** Redshift PDFs for galaxies in each of our three redshift samples, stacked in bins  $\Delta M_{UV} = 1$ . The legends give the number of galaxies in each stack, as well as the fraction of the redshift PDF at  $z > 4$  (denoted as P). Even in the worst case (which is for faint galaxies at  $z = 6$ )  $\lesssim 16.3\%$  of the sample could possibly be at lower redshift.

our requirement of detections in two bands (as well as the frequent detections in more than two bands for all but the highest redshift objects in the  $z = 8$  sample) makes a significant contamination by these sources unlikely. The most likely possible contaminants are faint red galaxies at  $z \leq 2$  (e.g., Dickinson et al. 2000). These galaxies can be too faint to be detected in our optical imaging, but their red SEDs yield detections in the WFC3/IR bands. Although faint sources that are very red will have a disfavored high-redshift solution with our current photometric selection, we have information on their likelihoods encoded in our redshift PDFs. Figure 5 shows the redshift PDFs of galaxies in each of our three highest redshift samples, stacked in magnitude bins of  $\Delta M = 1$  mag. At all redshifts and all magnitudes,  $\gtrsim 85\%$  of the redshift PDF is at  $z > 4$ , implying that there is not significant contamination by lower-redshift galaxies.

The position of the secondary redshift peak is consistent with the detection of a  $4000 \text{ \AA}$  break, rather than the Lyman break (at  $z = 6, 7,$  and  $8$ , this gives  $z_{\text{secondary}} = 1.1, 1.4,$  and  $1.7$ , respectively). At  $z = 8$ , the possible contamination from galaxies at  $z < 4$  is  $< 10.5\%$ , which is primarily due to the fact that at  $z = 8$ , a galaxy will have to be undetected in most of the filters we consider here (there is an additional  $\sim 8\%$  chance the true redshift is in the range  $4 < z < 6$ ). The worst case is for faint galaxies at  $z = 6$ , because  $z = 6$  galaxies are typically detected in all but two filters, although even here, the indicated contamination by  $z < 4$  galaxies is  $\lesssim 15\%$ .

### 3.8.3. Stacking Imaging

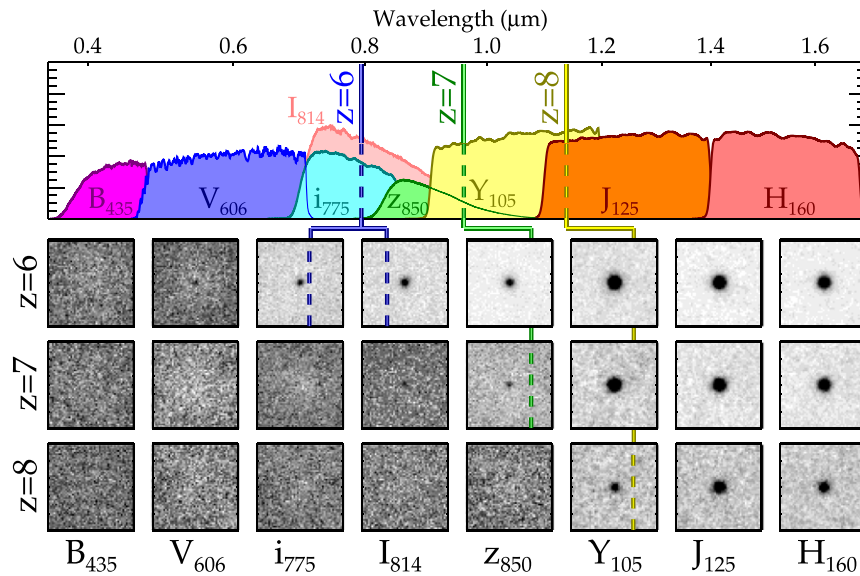
The limits from the previous subsection are likely upper limits on the contamination fraction. When fitting photometric redshifts to rule out all low-redshift solutions, the Lyman break needs to be detected at high significance, which is the case for only the brightest galaxies (e.g., at  $z = 6$ , the brightest bin has a contamination of  $< 2\%$ ). Additionally, these results are dependent on the templates used, which by definition do not account for unknown galaxy populations. We therefore consider two empirical tests of contamination. The first is to stack all galaxies in a sample to search for detections below the Lyman break; the results from this test for  $z = 6, 7,$  and  $8$  are

shown in Figure 6. As expected for galaxies at the expected redshifts, there is no visible signal in the  $B_{435}$ -band at  $z = 6$ ,  $i_{775}$ -band and blueward at  $z = 7$ , and  $I_{814}$ -band and blueward at  $z = 8$ . This confirms that the majority of the flux in our sample does not arise from lower-redshift sources.

### 3.8.4. Estimates from Dimmed Real Sources

As a final test, we estimated the contamination by artificially dimming real lower redshift sources in our catalog, to see if the increased photometric scatter allows them to be selected as high-redshift candidates. This empirical test is useful because it does not rely on known spectral templates to derive the contamination, although it does assume that the fainter objects that could potentially contaminate our sample have similar SEDs to the bright objects that we dim. We performed this exercise twice, once using the combined catalog of the GOODS-S and GOODS-N Deep fields, and once in the HUDF main field, to probe fainter magnitudes. In the GOODS Deep fields, we selected all real sources with  $21 < H_{160} < 24$  and  $z_{\text{phot}} < 3$ , and reduced their observed fluxes by a factor of 20. The same was done for sources drawn from the HUDF Main field, here extending the magnitude range to be  $21 < H_{160} < 26$ . The limits on these magnitudes were chosen to exclude any real high-redshift sources. We replaced the true flux uncertainties of these objects with flux uncertainties from a randomly drawn real source from the full catalog from a given field with a similar magnitude as the dimmed source. We then added scatter to the dimmed fluxes, perturbing them by a random amount drawn from a Gaussian distribution with a standard deviation equal to the flux uncertainties of the object. We included two realizations of the HUDF field to increase the number of dimmed objects.

The total number of sources in our artificially dimmed catalog was 4066 in the Deep fields and 1254 in the HUDF field. We measured the photometric redshifts of these sources with EAZY in an identical manner as on our real catalogs, and then applied our sample selection to this dimmed catalog. In the Deep fields, we found that a total of 149, 134, 54, 23, and 8 dimmed objects satisfied our  $z = 4, 5, 6, 7,$  and  $8$  selection criteria. Investigating the original (not dimmed or perturbed)



**Figure 6.** Top: filter transmission curves for the filter set used in this study ( $Y_{098}$ , which was used in the GOODS-S ERS field only, is not shown). The vertical lines denote the relative position of the Ly $\alpha$  break (rest 1216 Å) in a given filter for galaxies at the center of our three highest redshift bins. (Bottom) negative mage stacks of galaxies in our three highest redshift galaxy samples. If our sample had a significant fraction of lower-redshift interlopers, significant flux would be seen blueward of the break (e.g.,  $B_{435}$  at  $z = 6$ ,  $i_{775}$  and blueward at  $z = 7$ , and  $I_{814}$  and blueward at  $z = 8$ ). This is not observed at any redshift, thus we conclude that our sample does not contain a dominant population of low-redshift interlopers.

colors of these sources, we found that they are unsurprisingly red, with the bulk of sources having  $V - H > 2$  mag. Thus this parent population of red sources is responsible for the majority of the possible contamination. The contamination fraction in our high-redshift sample is then defined as

$$\mathcal{F} = \frac{N_{\text{dimmed,select}}}{N_{\text{dimmed,red}}} * \frac{N_{\text{total,red}}}{N_z} \quad (1)$$

where  $N_{\text{dimmed,select}}$  was the number of dimmed sources satisfying our high-redshift sample selection;  $N_{\text{dimmed,red}}$  was the total number of sources in the dimmed catalog with original colors of  $V - H < 2$ ;  $N_{\text{total,red}}$  was the number of sources in the full object catalog with  $25 < H < 27$ ,  $z_{\text{phot}} < 3$ , and  $V - H < 2$ ; and  $N_z$  was the number of true galaxy candidates in a given redshift bin. For example, at  $z = 6$ , where we found 54 dimmed galaxies that satisfied our selection criteria ( $N_{\text{dimmed,select}} = 54$ ),  $N_{\text{dimmed,red}} = 1023$ ,  $N_{\text{total,red}} = 695$ , and  $N_z = 322$ , giving an estimated contamination fraction  $\mathcal{F} = 11.4\%$ . Thus for sources with  $25 < H < 27$ , we found an estimated contamination fraction of  $\mathcal{F} = 4.5\%$ ,  $8.1\%$ ,  $11.4\%$ ,  $11.1\%$ , and  $16.0\%$  at  $z = 4, 5, 6, 7$ , and  $8$ , respectively. We performed the same exercise in the HUDF, for fainter sources with  $26 < H < 29$ , finding that 30, 21, 8, 8, and 0 sources satisfied our  $z = 4, 5, 6, 7$ , and  $8$  selection criteria, respectively, giving a contamination fraction of  $9.1\%$ ,  $11.6\%$ ,  $6.2\%$ ,  $14.7\%$ , and  $< 4.9\%$  at  $z = 4, 5, 6, 7$ , and  $8$ , respectively.

Broadly speaking, we estimate relatively small contamination fractions of  $\sim 5\%$ – $15\%$ , which is in line with the estimates from the stacked  $P(z)$  curves. As the bulk of contaminants appear to be red galaxies, it is interesting to compare to the space density of these potentially contaminating sources. This was recently estimated by Casey et al. (2014b), who found that dusty star-forming galaxies at  $z < 5$  will contaminate  $z > 5$  galaxy samples at a rate of  $< 1\%$ . This is much less than our

contamination estimates, thus we may have overestimated the contamination rate; although it may not be inconsistent once photometric scatter is applied to faint, red galaxies, making it easier for them to scatter into our sample. In any case, the expected contamination rate is quite small, therefore we do not reduce our observed number densities for the expected minimal contamination.

#### 4. COMPLETENESS SIMULATIONS

We performed an extensive set of simulations to estimate the effective volume for each source in our sample, accounting for both image incompleteness and selection effects. We inserted mock galaxies into the imaging data, repeating the same analysis for source detection, photometry, photometric redshift measurement, and sample selection as was done on the real data. We then compared the fraction of recovered and selected mock sources to the total number of input sources in a given bin of absolute magnitude and redshift to determine our completeness in that bin.

While the effective volumes are typically computed as a function of magnitude and redshift, other key factors in these simulations are the choices of galaxy size and color. At a constant magnitude, a very extended galaxy may not make it into our sample, because it may fall below our surface brightness sensitivity. Additionally, very red galaxies may not satisfy our selection criteria because red colors typically enhance the amplitude of the low-redshift solutions, particularly at low signal-to-noise levels. Thus the effective volume depends not only on magnitude and redshift, but also on the size and rest-frame UV color. To see what effect this has, we computed our completeness as a function of four properties: redshift, absolute magnitude, half-light radius, and rest-UV color, where we have parametrized the latter via the UV spectral slope  $\beta$  (Calzetti et al. 1994). A large number of simulated objects are needed to fill out this four-dimensional space; our completed simulations recovered  $\sim 5.4$  million out

of 7 million objects input across all of our fields (where the recovered objects were detected in our photometry catalogs; this number does not account for the photometric redshift selection, which we will discuss).

Our simulations were run separately on each of our 10 subfields, defined in Table 1. To ensure that the mock galaxies did not affect the background estimation, only a small number of galaxies were added during each simulation. To optimize the simulation runtime, the mock galaxies were added to cutouts from the full images. In the GOODS subfields (i.e., CANDELS Deep and Wide, and the ERS), 200 mock galaxies were added to a  $2000 \times 2000$  pixel ( $2' \times 2'$ ) region of the images, whereas for the single-pointing HUDF and HFF fields, 100 galaxies were added to a  $1000 \times 1000$  pixel region. Because the depth across our imaging data can vary, the position of the cutout varied during each simulation, such that when we combine all of our simulations, we average over any differences in the depth across a given field.

To determine the colors of the mock galaxies, we created distributions in redshift, dust attenuation (parameterized by  $E[B - V]$ ), stellar population age, and stellar metallicity. The redshift distribution was defined to be flat across  $3 < z < 9$ , such that we simulate objects well above and below the redshift ranges of interest. The dust attenuation  $E(B - V)$  was defined to have a Gaussian distribution with a mean of 0.1 and a  $\sigma = 0.15$  (with a minimum of zero). The age was defined as a log-normal distribution, with a peak near 10 Myr, and a tail extending out to the age of the universe at a given redshift. The metallicity distribution was also log-normal, with a peak of  $Z = 0.2Z_{\odot}$ , and a tail toward higher values. The exact values of these distributions are not crucial given our methodology (as opposed to a multivariate analysis, where the distributions are very important), because they combine to create a distribution of rest-frame UV slope,  $\beta$ . We crafted these distributions to provide a distribution of  $\beta$  encompassing the expected values for our real objects (e.g., Finkelstein et al. 2012b; Bouwens et al. 2014). We then used the updated (2007) stellar population models of Bruzual & Charlot (2003) to calculate the colors of a stellar population given the distributions above. To convert these colors into magnitudes, we assumed a distribution of H-band magnitudes designed to have many faint ( $H > 26$ ) galaxies (which is where we expect to become incomplete), and relatively few at bright magnitudes. To ensure there were enough bright galaxies to calculate a robust incompleteness, every 10th simulation used a flat distribution of H-band magnitudes of  $22 < H < 25$ . These H-band magnitudes were combined with the mock galaxy colors to generate magnitudes in each filter for a given field.

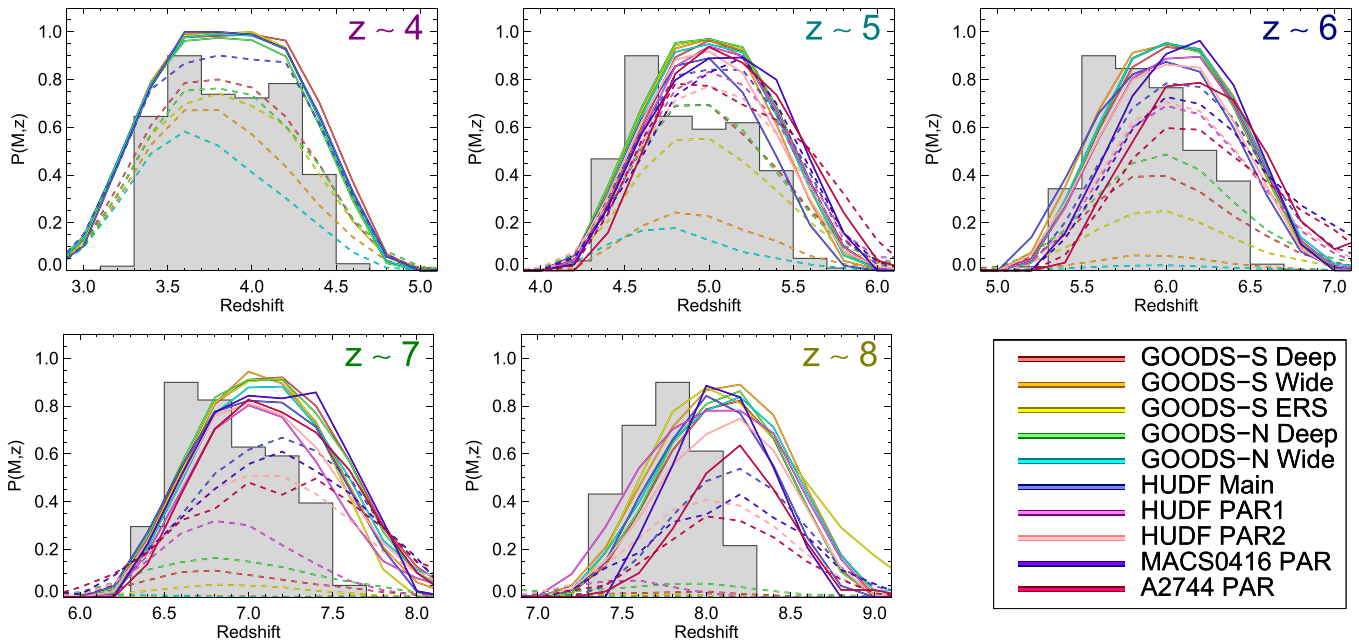
To generate the galaxy images themselves, we used the GALFIT software (Peng et al. 2002). We assumed a log-normal distribution of half-light radii, with a peak at 1-pixel and a high tail toward larger radii, giving an interquartile range of half-light radii of 1.4–4.9 pixels. This corresponds to  $\sim 0.4$ –1.6 kpc, spanning the range of the majority of resolved galaxies at  $z > 4$  (e.g., Oesch et al. 2010a; Grazian et al. 2012; Ono et al. 2013; Curtis-Lake et al. 2014). GALFIT also requires a Sérsic index ( $n$ ), axis ratio, and position angle; the Sérsic index was assumed to be a log-normal distribution at  $1 < n < 4$ , with the majority of the mock galaxies having disk-like morphologies ( $n < 2$ ); the axial ratio was also log-normal, with a peak at 0.8, and a tail toward lower values; and the position angle was a uniformly distributed random value between  $0^{\circ}$  and  $360^{\circ}$ .

GALFIT was then used to generate a  $101 \times 101$  pixel ( $6'' \times 6''$ ) stamp for a given mock galaxy, which was then added to the image at a random location. Because our data are PSF-matched to the H band, we had GALFIT convolve the mock galaxy images with the H-band PSF prior to adding them to the data for all filters.

Once the set of mock galaxies for a given simulation was added to the data, photometric catalogs were generated using Source Extractor in the exact same manner as was done on the data (i.e., using a weighted  $J+H$  image as the detection image). These catalogs were read in and combined, again in the same methodology as with the data, including aperture corrections (the exception here is that a correction for Galactic extinction was *not* applied in the simulation, as the simulated objects did not include Galactic extinction). The photometric catalog was then compared to the input catalog to generate the list of recovered objects (i.e., mock galaxies that were recovered by Source Extractor); an object was regarded as being recovered if it had a positional match within  $0''.2$  of one of the input mock galaxies. The recovered object catalogs were processed through EAZY to generate photometric redshifts, and then run through our SED-fitting routine to measure absolute UV magnitudes ( $M_{1500}$ ), stellar masses, and UV spectral slopes. These simulations were then repeated until a large sample of recovered galaxies was available, which was then compiled in a single database per field. The completeness was defined as the number of galaxies recovered versus the number of input galaxies, as a function of input absolute magnitude, redshift, half-light radius, and UV spectral slope  $\beta$ . Figure 7 shows the results from our simulations.

In our original simulations the recovered redshift was typically  $\sim 0.2$  lower than the input redshift, independent of magnitude. This is likely not a fault of our photometric redshift estimates, as Figure 2 shows that these agree well with existing spectroscopic redshifts for real galaxies. Rather, it is likely a mismatch between our simulated SEDs and those of the templates used in EAZY. Upon further investigation, we found that the cause of this offset was  $\text{Ly}\alpha$  emission in the mock galaxies. While  $\text{Ly}\alpha$  photons were attenuated by dust in the same manner as adjacent UV photons, we did not include any additional  $\text{Ly}\alpha$  attenuation due to, for example, geometric or kinematic effects. This led to very high  $\text{Ly}\alpha$  escape fractions, which were not matched in the templates. This high  $\text{Ly}\alpha$  emission reduced the amplitude of the photometrically measured Lyman break, resulting in a (slightly) lower photometric redshift. After reducing the amount of  $\text{Ly}\alpha$  flux to 25% of the intrinsic value, our photometric redshifts matched the input redshifts (which matches expectations for the global  $\text{Ly}\alpha$  escape fraction; e.g., Blanc et al. 2011; Hayes et al. 2011). Rather than rerun all of our completeness simulations, we elected to simply reduce the input redshift by 0.2 when interpreting our simulations, which corrects for this effect (this changes the distance modulus by  $< 0.1$  mag). The exception was the simulations for the HFF parallel fields, which were run after this effect was noticed. In those fields, the  $\text{Ly}\alpha$  flux of the input models was reduced to 25% of the intrinsic value, and no change was needed to the model redshift.

It is important to examine whether the choice of computing the completeness as a function of input properties affects our result. As mentioned in Section 2.3, we used the results from these simulations to correct for offsets in the recovered versus



**Figure 7.** Results of our completeness simulations, showing the probability that a given simulated source was recovered as a function of its input redshift. The solid lines denote sources with  $M_{1500} = -22$ , while the dashed lines denote  $M_{1500} = -19$ . These lines assume a half-light radius of  $r_h = 0''.18$  and  $\beta = -2.0$ . The background histogram shows the (normalized) distribution of best-fit photometric redshifts for the real galaxies in each redshift subsample. Although our selection criteria combined with the wavelengths probed by our filter set result in a completeness that peaks at close to  $z = 4, 5, 6, 7$ , or  $8$ , the evolving luminosity function as well as our sensitivity to bright galaxies results in our samples having mean redshifts that are slightly lower than the bin center, particularly in the higher redshift samples.

input magnitudes (i.e., to be sure the fluxes we use represent the total flux). Additionally, we examined whether there exist biases in the half-light radius or  $\beta$  measurements from the simulations. Recovered objects were typically measured to have a half-light radius  $\sim 0''.03$  (0.5 pixels) smaller than the input value, and were measured to be slightly redder ( $\Delta\beta \lesssim 0.1$ ). However, these corrections make effectively no change to the effective volumes derived from the simulations, and so were not applied.

In each redshift bin, the effective volume for galaxies in a given field was then calculated via

$$V_{\text{eff}}(M_{1500}, r_h, \beta) = \int \frac{dV}{dz} P(M_{1500}, z, r_h, \beta) dz \quad (2)$$

where  $dV/dz$  is the comoving volume element, and  $P(M_{1500}, z, r_h, \beta)$  is the result from our completeness simulations. The integral was done over  $\Delta z = 1$ , centered on the center of each redshift bin. In each field, we used a weighted mean of this three-dimensional effective volume  $V_{\text{eff}}(M_{1500}, r_h, \beta)$  to calculate  $V_{\text{eff}}(M_{1500})$ , where the weighting is based on the number of real objects in a magnitude bin with a given value of  $r_h$  and  $\beta$ , as

$$V_{\text{eff}}(M_{1500}) = \frac{\sum_{r_h} \sum_{\beta} V_{\text{eff}}(M_{1500}, r_h, \beta) N(M_{1500}, r_h, \beta)}{\sum_{r_h} \sum_{\beta} N(M_{1500}, r_h, \beta)}. \quad (3)$$

This assumes that the completeness corrections estimated using our observed size and color distributions are similar to what we obtained if we could measure the true sizes and colors, motivated by our measurement of minimal size and color biases when comparing the input to recovered values.

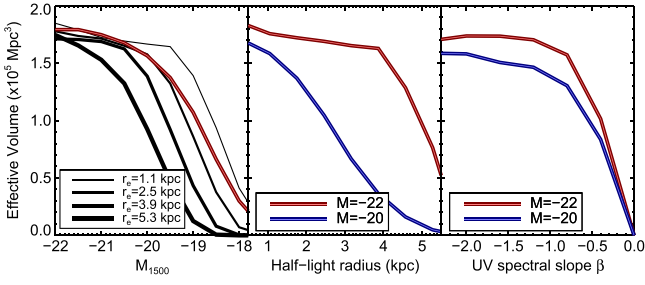
This weighted volume is the most representative of the true volume we are sensitive to, as we explicitly account for the

incompleteness as a function of size and color. Figure 8 highlights the dependence of the effective volume on these quantities for galaxies in our  $z = 6$  sample in the GOODS-N Deep field. The effective volume has a strong dependence on the surface brightness of galaxies, as the volume drops steeply both for larger sizes and fainter magnitudes. The central and right panels highlight that while the effective volume (and thus sample completeness) is sensitive to both size and color, the color has a relatively minor role. We remain sensitive to fairly red galaxies ( $\beta = -1$ ), similar to previous results from Bouwens et al. (2012). Although the effective volume has a strong dependence on size, the relatively small sizes of galaxies in our sample yields a volume similar to that obtained when assuming a constant small size. Thus, although our volumes are the most accurate, had we assumed a fixed effective radius of, for example,  $r_e = 1$  kpc, our results would not change significantly. This is consistent with the conclusions of Grazian et al. (2012) who, when accounting for the size–luminosity relation when deriving the  $z = 7$  luminosity function, found similar results to previous studies that had neglected the size–luminosity relation. Our final effective volumes are shown in Figure 9.

## 5. THE LUMINOSITY FUNCTION

### 5.1. Parametric Approach

Now that we have a final galaxy sample with measured values of  $M_{1500}$ , as well as the effective volumes from our completeness simulations in the previous section, we can proceed to measure the rest-frame UV luminosity function at  $z = 4, 5, 6, 7$ , and  $8$ . We calculate the luminosity function in two ways: a parametric version assuming that the luminosity function takes the form of a Schechter (1976) function, and a non-parametric stepwise maximum likelihood (SWML) calculation.



**Figure 8.** Effective volume per unit area of our survey for high-redshift galaxies in each of our redshift bins. Here we divide out the area of a given subfield, such that one can easily compare the completeness as a function of magnitude of the various fields. The solid lines give way to dashed lines when the volume per unit area falls below 50% of the maximum value. For the luminosity function, we only consider magnitude bins in each field that are brighter than these 50% completeness points, to avoid having data dominated by incompleteness corrections. The ERS has a different  $Y$ -band filter ( $Y_{098}$ ), which gives a better spectral resolution around  $1 \mu\text{m}$ , likely responsible for the increased selection efficiency at  $z = 8$  in that field.

The fitting of a Schechter function is well motivated, as it successfully matches the observed rest-UV luminosity functions at lower redshifts (e.g., Bouwens et al. 2006; Reddy & Steidel 2009). This function is characterized by a power law at the faint end with slope  $\alpha$ , and an exponential cutoff at the bright end, transitioning between the two regimes at the characteristic magnitude  $M^*$ . The parameter  $\varphi^*$  sets the normalization of this function. The number density at a given magnitude is then given by

$$\varphi(M) = 0.4 \ln(10) \varphi^* 10^{-0.4(M-M^*)(\alpha+1)} e^{-10^{-0.4(M-M^*)}}. \quad (4)$$

For the measurement of the luminosity function assuming a Schechter functional form, we calculated the likelihood that the number of observed galaxies in a given magnitude bin is equal to that for an assumed value of the Schechter parameters  $M^*$

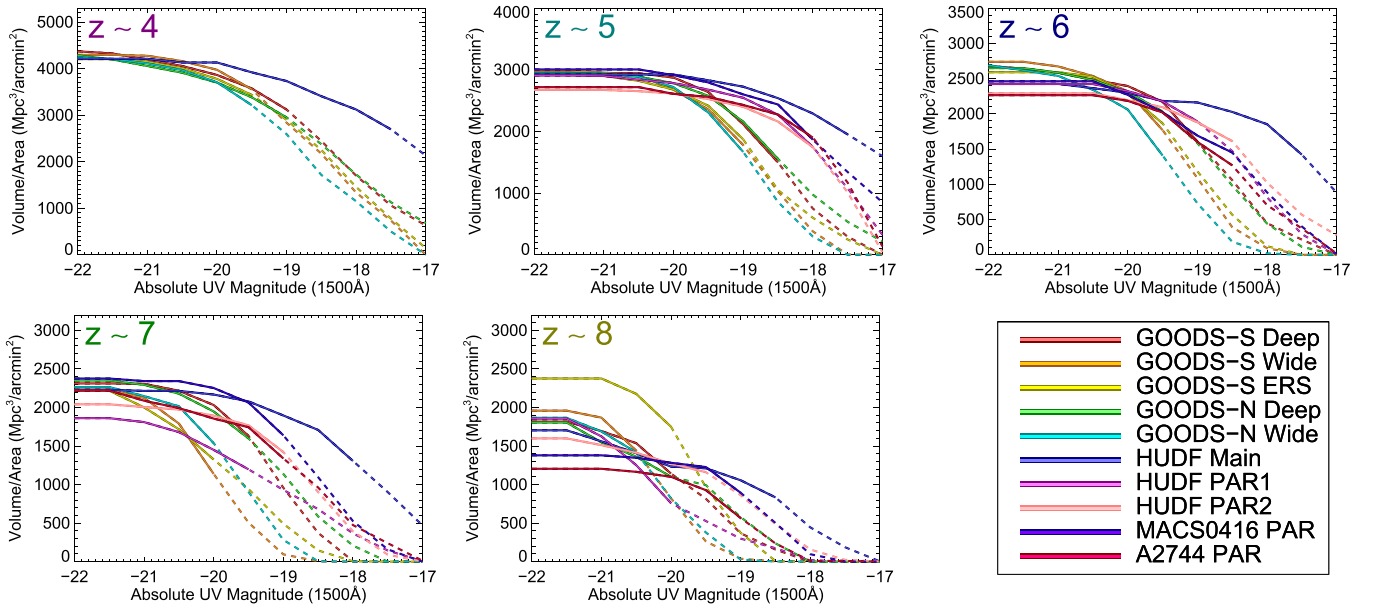
and  $\alpha$ . Rather than performing a grid-based search, we performed a Markov Chain Monte Carlo (MCMC) search algorithm to better span the parameter space, as well as to better characterize the uncertainties on the Schechter parameters. We performed this calculation in bins of absolute magnitude with  $\Delta M = 0.5 \text{ mag}$ , ranging from  $-24 \leq M_{1500} \leq -17$ . At the bright end we are in the limit of small numbers, therefore we model the probability distribution as a Poissonian distribution (e.g., Cash 1979; Ryan et al. 2011), with:

$$C^2(\varphi) = -2 \ln \mathcal{L}(\varphi) \quad (5)$$

$$C^2(\varphi) = -2 \sum_i \sum_j N_{j,\text{obs}} \ln(N_{j,\text{model}}) - N_{j,\text{model}} - \ln(N_{j,\text{obs}}!) \quad (6)$$

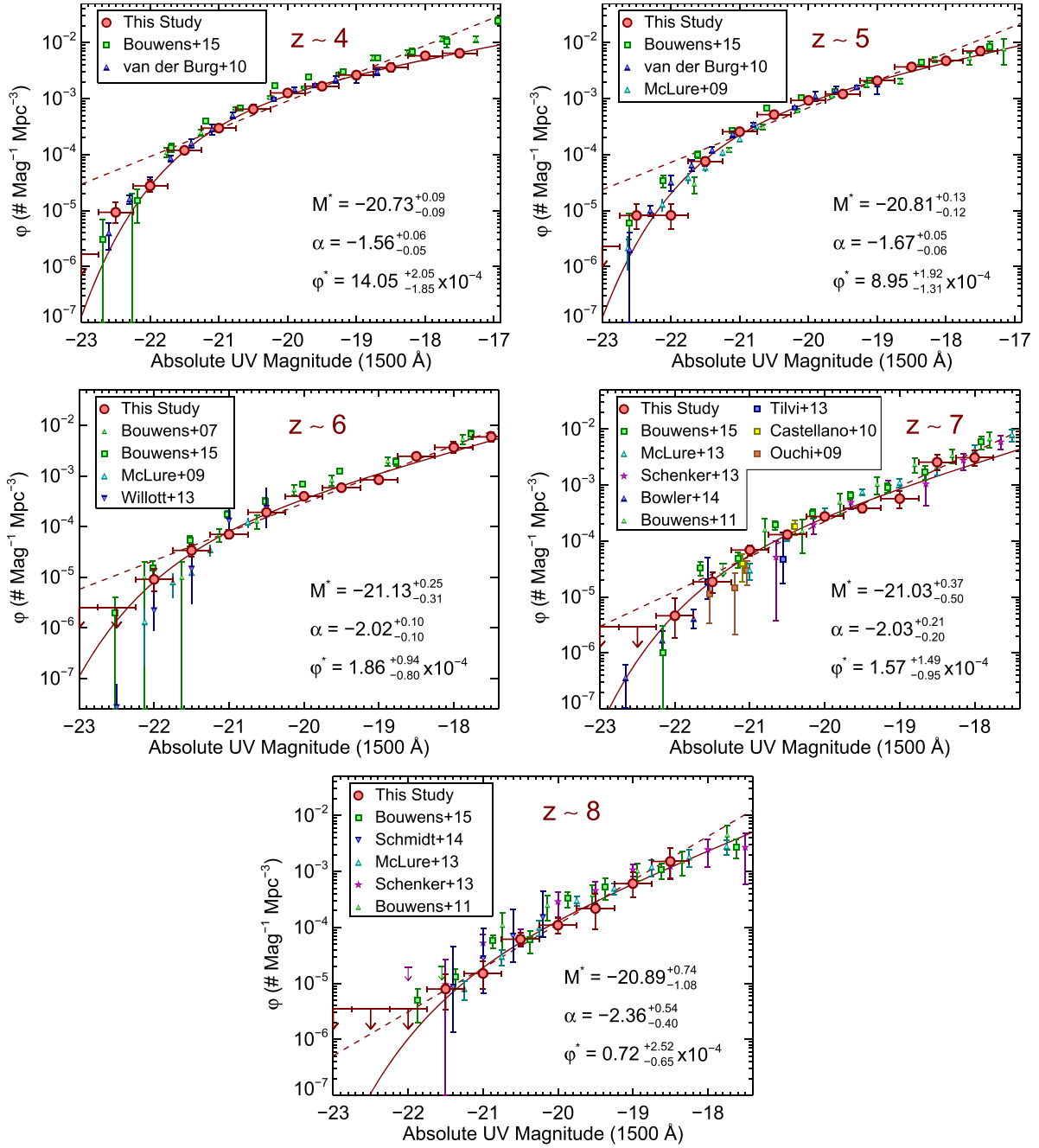
where  $\mathcal{L}(\varphi)$  is the likelihood that the expected number of galaxies ( $N_{\text{model}}$ ) matches that observed ( $N_{\text{obs}}$ ) for a given value of  $M^*$  and  $\alpha$ , and  $C^2$  is the goodness-of-fit statistic. The subscripts  $i$  and  $j$  represent the subfields and magnitude bins, respectively. The final goodness-of-fit is the sum over all fields and magnitudes in a given redshift bin. We use the effective volume results for a given redshift, magnitude bin, and field to convert from the model number density to the expected number, calculating  $\varphi^*$  as the normalization, such that the total expected number of galaxies over all magnitude bins matches the total number of observed galaxies.

For each magnitude bin, we performed 10 independent MCMC chains utilizing a Metropolis–Hastings algorithm, each of  $10^5$  steps, building a distribution of  $M^*$ ,  $\alpha$ , and  $\varphi^*$  values for each field. During each step of the chain, the likelihood of a given model was computed for each of our observed fields, and then added together to compute the likelihood for the sample as a whole (we also recorded the individual field values, see



**Figure 9.** (Left) effective volume as a function of UV absolute magnitude for galaxies in our  $z = 6$  sample in the GOODS-N Deep field. The red line shows the mean effective volume for this field, weighted by the number of galaxies at a given radius and UV color. The black lines show how the effective volume changes as a function of effective radius ( $r_e$ ) for a fixed UV color ( $\beta = -2$ ). Our weighted mean volume is similar to the effective volume, assuming  $r_e = 2.5 \text{ kpc}$  for bright galaxies and  $r_e = 1.1 \text{ kpc}$  for faint galaxies. (Middle) the dependence of the effective volume on effective radius in two magnitude bins. At fainter magnitudes, the effective volume drops steeply with increasing size, as the surface brightness drops below detectable levels. (Right) same as middle, except here showing the dependence on UV color. The dependence on color is much weaker than that on size, as we remain sensitive to galaxies until  $\beta$  becomes redder than  $-1$ , which is much redder than the colors of observed high-redshift galaxies (e.g., Finkelstein et al. 2012b; Bouwens et al. 2014).





**Figure 10.** Rest-frame UV luminosity functions for our  $z = 4$ – $8$  galaxy samples. The large red circles denote our stepwise maximum likelihood luminosity function, while the solid red line denotes our best-fitting Schechter function, with the best-fit values given by the inset text. We do not use data below the determined 50% completeness level in each field. As the HUDF is our deepest field, the magnitude of our last data point denotes the 50% completeness limit in the HUDF. The dashed lines show the best-fit single power law at each redshift. We also show several luminosity functions from the literature, as indicated in the legends.

Section 6.5). Prior to each recorded chain, we performed a burn-in run with a number of steps equal to 10% of the number of steps in each chain. The starting point for the burn is a brute-force  $\chi^2$  fit of a grid of  $\alpha$  and  $M^*$  values to our data. At the end of the burn, the final values of the parameters from the last step were then the starting points for each chain. The rest of the burn-in results were not recorded. During each step, new values of  $M^*$  and  $\alpha$  were chosen from a random Gaussian distribution, with the Gaussian width tuned to generate an approximate acceptance rate of 23%. During each step  $\phi^*$  was calculated as the normalization. If the difference between the likelihood of the model for the current step exceeds that from the previous

step by more than a randomly drawn value (i.e.,  $\equiv 2 \ln(n)$ ; where  $n$  is a uniform random number between zero and unity), then the current values of the Schechter function parameters were recorded. If not, the chain reverted to the value from the previous step.

By running 10 independent chains, we mitigate against being trapped by local minima in the parameter space. Our final result links these 10 chains together, giving a distribution of  $10^6$  values of the Schechter function parameters at each redshift. The results were visually inspected to confirm that the chains reached convergence. For each Schechter function parameter, the best-fit values were taken to be the median of the

**Table 4**  
Schechter Function Fits to the Luminosity Function

Redshift	$M_{UV}^*$	$\alpha$	$\varphi^*$ ( $\text{Mpc}^{-3}$ )
4	$-20.73^{+0.09}_{-0.09}$	$-1.56^{+0.06}_{-0.05}$	$(14.1^{+2.05}_{-1.85}) \times 10^{-4}$
5	$-20.81^{+0.13}_{-0.12}$	$-1.67^{+0.05}_{-0.06}$	$(8.95^{+1.92}_{-1.31}) \times 10^{-4}$
6	$-21.13^{+0.25}_{-0.31}$	$-2.02^{+0.10}_{-0.10}$	$(1.86^{+0.94}_{-0.80}) \times 10^{-4}$
7	$-21.03^{+0.37}_{-0.50}$	$-2.03^{+0.21}_{-0.20}$	$(1.57^{+1.49}_{-0.95}) \times 10^{-4}$
8	$-20.89^{+0.74}_{-1.08}$	$-2.36^{+0.54}_{-0.40}$	$(0.72^{+2.52}_{-0.65}) \times 10^{-4}$

**Note.** The final values for each parameter are the median of the parameter distribution from the MCMC analysis. The quoted errors represent the 68% confidence range on each parameter.

distribution, with the uncertainty being the central 68% of the distribution. These results are given in Table 4. For the  $z = 8$  Schechter function fit, we imposed a top-hat prior, forcing  $M_{UV}^*$  to be fainter than  $-23$ . Without this prior, the fit preferred a much brighter value of  $M_{UV}^*$ , such that the observed data points all lay on the faint-end slope (i.e., a single power law). We discuss the implications of this in Section 6.6.

Although we computed the volumes down to very faint magnitudes, we would be highly incomplete if we included these faint galaxies and calculated the luminosity function down to  $M_{1500} = -17$  or fainter (Figure 9). In practice, it is our deepest field (i.e., the HUDF) that determines how faint we can constrain the luminosity function. The HUDF drops below 50% completeness at magnitudes fainter than  $M_{1500} \sim -17.5$  at  $z = 4, 5,$  and  $6$ ;  $-18$  at  $z = 7$ ; and  $-18.5$  at  $z = 8$ . Thus, in our calculation of the luminosity function, we only include a given field’s contribution at a given magnitude if it is above the 50% completeness limit for that magnitude and redshift. Extending the analysis fainter will give results that are dominated by the incompleteness correction.

As shown in Figure 9, while the volume per unit area for the different fields is very tight at  $z = 4$ , there is a progressively larger scatter apparent when moving toward higher redshift, representing a systematic uncertainty in the effective volume calculation. One likely culprit is the fact that the volumes depend on the distribution of the sizes and colors of objects in a given field. For fields with few sources (i.e., the smallest fields at the highest redshifts), there may only be a single object in a given magnitude bin. To mitigate significant variances in the effective volume at the bright end, where the numbers of sources are small, we set the effective volume in a given redshift bin and field in bright bins with less than three objects equal to the value in the brightest bin with more than three objects (i.e., if there are no magnitude bins with more than three objects at  $M < -21$ , the effective volumes for all brighter bins are set equal to the value at  $M = -21$ ). This change has no discernable effect on our luminosity function results, as this is well above the 90% completeness limit for any of our fields and is thus only done to keep small numbers of galaxies from significantly affecting the volumes.

Another possible issue is the source density of simulated objects. In the smaller fields (i.e., HUDF, HUDF parallels, and HFF parallels) we input sources with twice the surface density as in the larger fields to speed up the computing time. As some sources (just like real galaxies) will inevitably fall on top of real sources, and thus not be recovered, an increased source density could result in a (slightly) lower completeness. This is just what

is observed in these fields, as shown in Figures 7 and 9. To account for this uncertainty, we measured the spread in volume per unit area in each field at  $M_{1500} = -21$  at each redshift, which we found to be  $\sim 1.5\%$ ,  $3.8\%$ ,  $6.2\%$ ,  $7.8\%$ , and  $13\%$  at  $z = 4, 5, 6, 7,$  and  $8$ , respectively. At each step in the MCMC chain, we perturbed the effective volume by this amount to account for this systematic uncertainty in our luminosity function results.

## 5.2. Non-parametric Approach

We also examined a non-parametric approach to studying evolution in the luminosity function. This is particularly warranted at very high redshift, where the effects responsible for suppressing the bright end of the luminosity function and causing the exponential decline in number density (e.g., AGN feedback or dust attenuation) may be less relevant. We thus calculated the SWML luminosity function, which is essentially the number density at a given magnitude, free from assumptions about the functional form of number density with magnitude. We also calculated the SWML luminosity function using an MCMC sampler. In this case, since the number densities in the magnitude bins are not linked by an overarching function, we calculated the number density in each magnitude bin independently.

For each magnitude bin and for each field, the likelihood was calculated using Equations (5) and (6) that a given randomly drawn value of  $\varphi(M)$  will give the observed number of galaxies. The actual recorded value of  $\varphi(M)$  is that which maximizes the likelihood. While in practice, this yields very similar results as one would get by simply taking the observed number and dividing by the effective volume (consistent within a few percent for bins with more than a few galaxies), our approach has two advantages. First, in the limits where numbers are small, this approach is more accurate because it properly accounts for the Poissonian likelihood. Second, this approach generates a full probability distribution for the number densities in each magnitude bin, allowing for the derivation of accurate asymmetric uncertainties. Our SWML luminosity function determinations and best-fit Schechter functions are given in Table 5 and shown in Figure 10.

## 6. LUMINOSITY FUNCTION INTERPRETATION

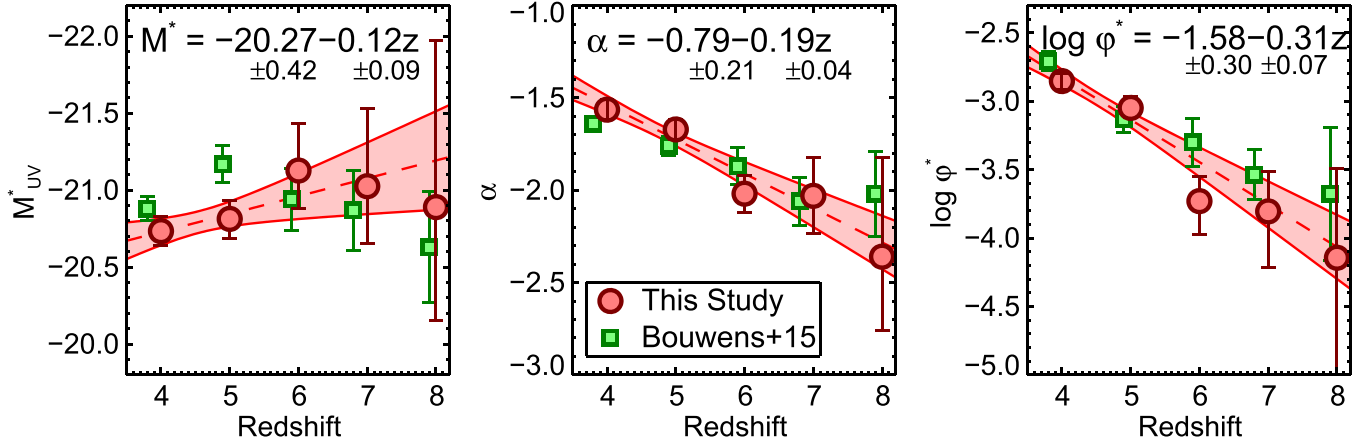
### 6.1. Evolution

As shown in Figure 10, the qualitative shape of the SWML luminosity functions at all redshifts we consider here are similar, in that bright galaxies are rare and faint galaxies are relatively common. Additionally, when examining the Schechter fits (solid line), we see that they are consistent with the SWML determinations. Surprisingly, the best-fit Schechter function parameters (Table 4) show little evolution in  $M_{UV}^*$ . However, from  $z = 4$  to  $8$ , the uncertainty on  $M_{UV}^*$  gets progressively larger, to  $0.4$  ( $0.9$ ) mag at  $z = 7$  ( $8$ ). This is easy to understand, as at all redshifts, our data set contains galaxies in only 1–2 bins brightward of  $M_{UV}^*$ . Ideally, one would prefer to have multiple bins in magnitude on either side of  $M_{UV}^*$  to obtain robust constraints. However, that requires a larger volume than we consider in this analysis. In Figure 11, we fit the evolution of  $M_{UV}^*$  with redshift with a linear function, using our results at  $z = 4$ – $8$ . We find that  $dM^*/dz = -0.12 \pm 0.09$ ; thus, our data do not support a significant evolution of  $M_{UV}^*$  with redshift.

**Table 5**  
Rest-frame Ultraviolet Luminosity Function: Stepwise Maximum Likelihood Method

$M_{1500}$	$\varphi(z \approx 4)$ ( $10^{-3} \text{ Mpc}^{-3} \text{ mag}^{-1}$ )	$\varphi(z \approx 5)$ ( $10^{-3} \text{ Mpc}^{-3} \text{ mag}^{-1}$ )	$\varphi(z \approx 6)$ ( $10^{-3} \text{ Mpc}^{-3} \text{ mag}^{-1}$ )	$\varphi(z \approx 7)$ ( $10^{-3} \text{ Mpc}^{-3} \text{ mag}^{-1}$ )	$\varphi(z \approx 8)$ ( $10^{-3} \text{ Mpc}^{-3} \text{ mag}^{-1}$ )
-23.0	<0.0016	<0.0023	<0.0025	<0.0029	<0.0035
-22.5	$0.0093^{+0.0045}_{-0.0033}$	$0.0082^{+0.0050}_{-0.0035}$	<0.0025	<0.0029	<0.0035
-22.0	$0.0276^{+0.0074}_{-0.0062}$	$0.0082^{+0.0051}_{-0.0036}$	$0.0091^{+0.0057}_{-0.0039}$	$0.0046^{+0.0049}_{-0.0028}$	<0.0035
-21.5	$0.1192^{+0.0145}_{-0.0132}$	$0.0758^{+0.0137}_{-0.0125}$	$0.0338^{+0.0105}_{-0.0085}$	$0.0187^{+0.0085}_{-0.0067}$	$0.0079^{+0.0068}_{-0.0046}$
-21.0	$0.2968^{+0.0230}_{-0.0219}$	$0.2564^{+0.0255}_{-0.0240}$	$0.0703^{+0.0148}_{-0.0128}$	$0.0690^{+0.0156}_{-0.0144}$	$0.0150^{+0.0094}_{-0.0070}$
-20.5	$0.6491^{+0.0361}_{-0.0347}$	$0.5181^{+0.0365}_{-0.0338}$	$0.1910^{+0.0249}_{-0.0229}$	$0.1301^{+0.0239}_{-0.0200}$	$0.0615^{+0.0197}_{-0.0165}$
-20.0	$1.2637^{+0.0494}_{-0.0474}$	$0.9315^{+0.0477}_{-0.0482}$	$0.3970^{+0.0394}_{-0.0357}$	$0.2742^{+0.0379}_{-0.0329}$	$0.1097^{+0.0356}_{-0.0309}$
-19.5	$1.6645^{+0.0630}_{-0.0618}$	$1.2086^{+0.0488}_{-0.0666}$	$0.5858^{+0.0527}_{-0.0437}$	$0.3848^{+0.0633}_{-0.0586}$	$0.2174^{+0.1805}_{-0.1250}$
-19.0	$2.6392^{+0.1192}_{-0.1165}$	$2.0874^{+0.1212}_{-0.1147}$	$0.8375^{+0.0916}_{-0.0824}$	$0.5699^{+0.2229}_{-0.1817}$	$0.6073^{+0.3501}_{-0.2616}$
-18.5	$3.6169^{+0.6799}_{-0.6091}$	$3.6886^{+0.3864}_{-0.3725}$	$2.4450^{+0.3887}_{-0.3515}$	$2.5650^{+0.8735}_{-0.7161}$	$1.5110^{+1.0726}_{-0.7718}$
-18.0	$5.8343^{+0.8836}_{-0.8204}$	$4.7361^{+0.4823}_{-0.4413}$	$3.6662^{+1.0076}_{-0.8401}$	$3.0780^{+1.0837}_{-0.8845}$	...
-17.5	$6.4858^{+1.0166}_{-0.9467}$	$7.0842^{+1.2829}_{-1.1364}$	$5.9126^{+1.4481}_{-1.2338}$	...	...

**Note.** Magnitude bins with zero objects are shown as upper limits, calculated as  $1/V_{\text{eff}}/\Delta M$  in that magnitude bin for that redshift.



**Figure 11.** Evolution of the Schechter function parameters. Red circles show results from this study; green squares show results from Bouwens et al. (2015). Dashed lines show the best-fit evolution as a linear function of  $z$ , indicated in each panel; red shaded regions show the 68% confidence range of the linear form. Contrary to previous studies, we find no significant evolution in  $M_{\text{UV}}^*$ . We find significant ( $>4\sigma$ ) evolution in  $\alpha$ , from steeper slopes at higher redshift to shallower slopes at lower redshift, and in the characteristic number density  $\varphi^*$ , which evolves to higher values by a factor of  $20\times$  from  $z = 8$  to 4.

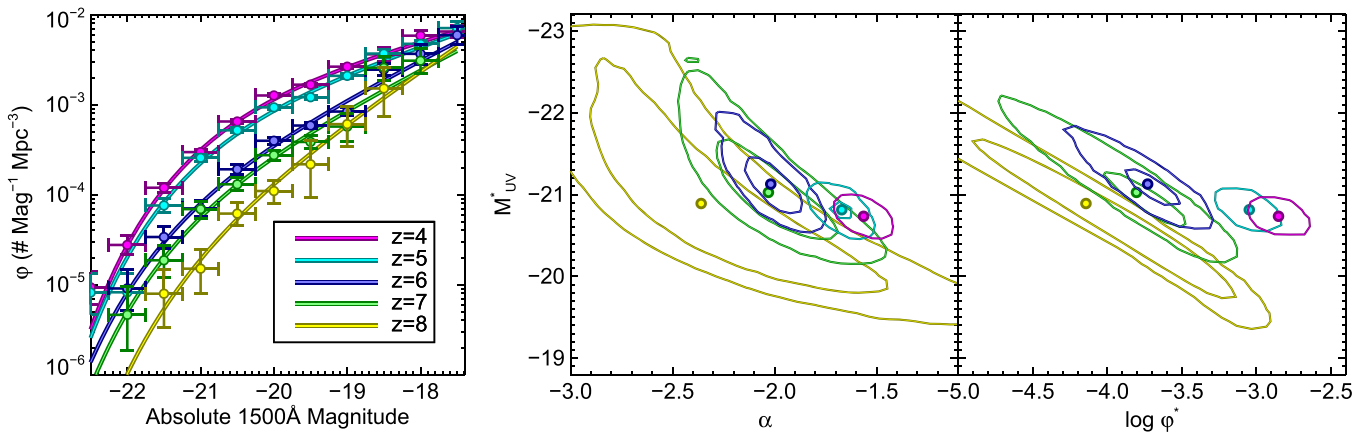
We also fit similar functions to see if we detect evolution in  $\alpha$  and  $\varphi^*$ . As shown in Figure 11, we do see significant evolution in the faint-end slope  $\alpha$ , with it becoming steeper at higher redshift, as  $d\alpha/dz = -0.19 \pm 0.04$  ( $4.8\sigma$  significance). We see a similar significance in the evolution of the characteristic number density  $\varphi^*$ , which evolves as  $d \log \varphi^*/dz = -0.31 \pm 0.07$  ( $4.4\sigma$  significance). Thus, while  $M_{\text{UV}}^*$  does not significantly evolve with redshift from  $z = 4$  to 8, both  $\alpha$  and  $\varphi^*$  do, in that the number density decreases and the faint-end slope becomes steeper with increasing redshift. In particular, this decline in characteristic number density is by a factor of  $\sim 20$  over a period of time of less than 1 Gyr. Although the steepening of the faint end is consistent with previous studies (e.g., Bouwens et al. 2012), the un-evolving  $M_{\text{UV}}^*$  and strong number density evolution are the opposite of what was presented in the literature just a year ago (e.g., Bouwens et al. 2007, 2011b; McLure et al. 2013). This updated evolutionary picture will be crucial when projecting number counts for future *HST* and *JWST* surveys.

In Figure 12, we show our determinations of the luminosity functions together at all five redshifts, along with the joint

confidence contours on  $M_{\text{UV}}^*$ ,  $\alpha$ , and  $\varphi^*$ . It is apparent that there is significant evolution in the luminosity function, with a drop in number density from  $z = 4$  to 8, as well as a gradual steepening of the faint-end slope. The apparent non-evolution of the characteristic magnitude is visible as the roughly constant magnitude of the “knee” of the luminosity function.

## 6.2. Impact of Magnitude Uncertainties

By definition, our method of computing the luminosity function is dependent on magnitude binning, because we compare the observed number to that expected based on a given model in magnitude bins of width 0.5 mag. While galaxies close to one side of a magnitude bin have the potential to scatter to another bin, the typical uncertainties on the UV absolute magnitudes of galaxies in our sample are  $\sim 0.2$  mag. Additionally, galaxies can shift both ways; thus, while one galaxy moves out of a bin, another may move in (although this effect will not be symmetric given the shape of the luminosity function). In our results in the previous section, we assumed that magnitude scatter does not significantly impact our results.

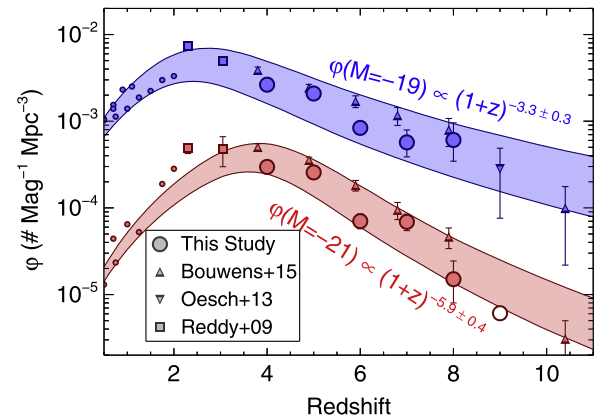


**Figure 12.** (Left) evolution of the UV luminosity function from  $z = 4$  to 8, where the circles and lines denote our stepwise and Schechter-parameterized luminosity functions, respectively. (Center) contours of covariance between  $\alpha$  and  $M_{UV}^*$  at  $z = 4, 5, 6, 7,$  and  $8$ . The contours denote the 68% and 95% confidence levels, while the small circles show the best-fitting values. (Right) contours of covariance between  $\varphi^*$  and  $M_{UV}^*$  at  $z = 4, 5, 6, 7,$  and  $8$ .

To investigate the impact of this assumption, we performed another iteration of MCMC fitting to our data, here allowing galaxies to scatter between magnitude bins. At each step in the MCMC chain, a new value of  $M_{1500}$  was drawn for each galaxy from the 100 SED-fitting Monte Carlo simulation results. The spread in these values encompassed both the photometric scatter in the observed filters and the uncertainty in the photometric redshift (see Section 3.7). To compare to our fiducial luminosity function values, we recorded both the median Schechter parameter results and the median number density in each magnitude bin, because, unlike our fiducial MCMC run, these varied during each step as the magnitudes changed. At all redshifts, our fiducial values of the stepwise luminosity functions are consistent with these “magnitude-scatter” values within 10% at  $M \geq -21.5$ , and typically within 2%–3%. The sole exception is in the brightest bin ( $-22$  at  $z = 4$ – $6$ , and  $-21.5$  at  $z = 7$  and  $8$ ), where our fiducial number density values are higher by  $\sim 15\%$ – $20\%$  (60% at  $z = 7$ , where there is only a single galaxy in this bin). We examined the Schechter fit to see whether this bright-end difference affects our results. Values of both  $M_{UV}^*$  and  $\alpha$  derived when allowing galaxies to shift between bins are consistent with our fiducial values within 0.1 mag and  $< 3\%$ , respectively. We conclude that the relatively small ( $\sim 20\%$ ) uncertainties in the absolute magnitudes of our galaxies do not have a significant impact on our luminosity function results.

### 6.3. Non-parametric Evolution

Given that our results show that the Schechter functional parameters may not be a robust method of tracking galaxy evolution (e.g., a non-evolving value of  $M_{UV}^*$  does not mean that the galaxy populations are not evolving), we examine the evolution in a non-parametric way. In Figure 13 we show the evolution of the stepwise luminosity function, plotting the number density corresponding to galaxies at  $M_{UV} = -21$  and  $-19$  versus redshift. From  $z = 8$  to 4, the abundance of brighter galaxies increases faster than faint galaxies. This trend halts at  $z = 4$ , where bright galaxies have an approximately constant abundance down to  $z = 2$ , and then turn over. Faint galaxies, however, continue increasing in abundance down to  $z = 2$ , where they also turn over. This figure highlights the phenomenon of downsizing, where bright/large galaxies grow faster at early times (e.g., Cowie et al. 1996, see also Lundgren

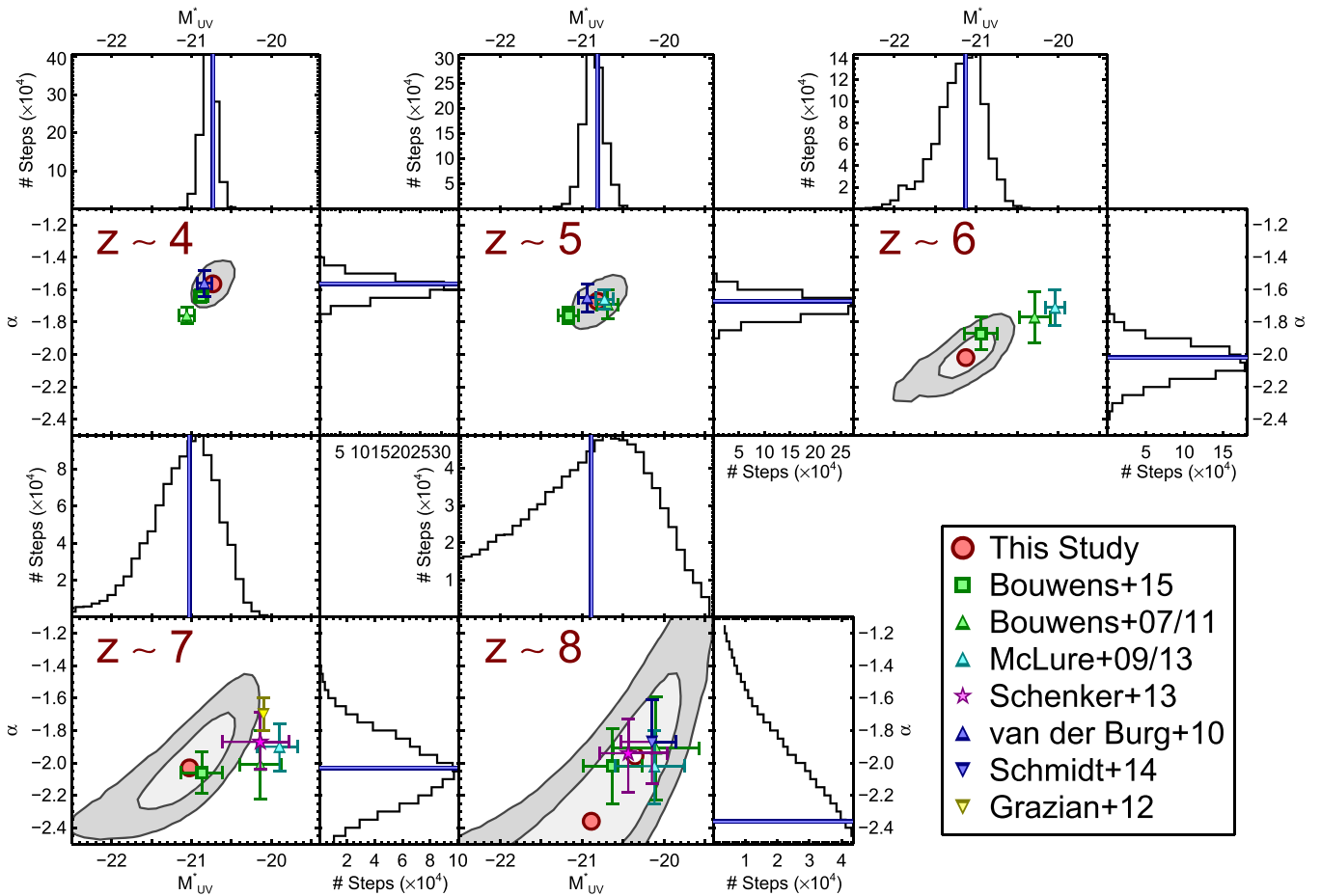


**Figure 13.** Number densities of bright ( $M_{UV} = -21$ ) and faint ( $M_{UV} = -19$ ) galaxies at a variety of redshifts. Our data are shown as large circles, and we also show results at high redshift from Bouwens et al. (2015), Oesch et al. (2013), and Reddy & Steidel (2009). At low redshift, we show results as small circles from Arnouts et al. (2005), Oesch et al. (2010b), and Cucciati et al. (2012). We fit the trend of  $\varphi$  with redshift, separately for our two magnitude bins, with the function given in Equation (7). The shaded regions show the 68% confidence ranges for each of the fits. The value of the slope of this function at high redshift is significantly steeper for bright galaxies than for faint galaxies, showing that from  $z = 8$  to 4, bright galaxies become more abundant at a faster rate than faint galaxies. This trend reverses at  $z = 4$ , where bright galaxies stop becoming more common. Bright galaxies peak in number density at  $z = 3.1$ – $3.2$ , sooner than faint galaxies, which peak at  $z = 2.4$ – $2.7$  (68% C. L.). At  $z < 2$ , the abundances of both populations plummet, in line with the evolution of the cosmic star-formation rate density.

et al. 2014). This is different from the expectation one would get simply from examining Schechter fits, because the luminosity functions don’t evolve much over the range  $2 < z < 4$  (e.g., Reddy & Steidel 2009). Given that the trends here mimic the evolution of the cosmic SFR density, we fit the function provided by Madau & Dickinson (2014) to our data for both number densities, given by

$$\varphi(z) = A \frac{(1+z)^\alpha}{1 + [(1+z)/B]^\gamma} \text{mag}^{-1} \text{Mpc}^{-3}. \quad (7)$$

The evolution with redshift is thus proportional to  $(1+z)^\alpha$  at low redshift, and  $(1+z)^{\alpha-\gamma}$  at high redshift, with  $A$  and  $B$  as dimensionless coefficients. Fitting all of the data, including the literature data shown in Figure 13, in this way, we confirm that



**Figure 14.** Confidence contours on our measured value of the faint-end slope  $\alpha$  and the characteristic magnitude  $M_{UV}^*$  at  $z = 4, 5, 6, 7,$  and  $8$ , with the light and dark shaded regions denoting 68% and 95% confidence, respectively. The large red circles represent our fiducial best-fit luminosity function parameters, while the other colored symbols denote results from previous studies, using the same symbols as in Figure 10 (with the addition of the results from Grazian et al. 2012, shown as the yellow triangle in the  $z = 7$  panel, who fit  $\alpha$  keeping  $M_{UV}^*$  fixed to  $-20.14$ ). In the  $z = 8$  panel, we also show our best-fit result when fixing  $\alpha \geq -2.3$  as the white-filled red circle. The histograms to the top and side of each contour plot show the number of MCMC steps when a given value of  $M_{UV}^*$  or  $\alpha$  was recorded, with the median value shown by the blue line.

at  $z > 3$ , bright galaxies change in abundance faster, as  $(1 + z)^{-5.9 \pm 0.4}$ , than faint galaxies, which go as  $(1 + z)^{-3.3 \pm 0.3}$ .

Another interesting aspect is to compare the trends observed to our predicted abundance of bright  $z = 9$  galaxies (see Section 9). The trend observed here slightly overestimates our predicted  $z = 9$  abundance, although if we assume the uncertainties on our  $z = 8$  number density applies to  $z = 9$ , our trend is consistent with this prediction. In any case, this trend of abundance with redshift lends more weight to our expectation of a significant abundance of bright  $z = 9$  galaxies. This figure neglects the impact of dust attenuation, because we are only looking at the observed UV magnitudes. The dust attenuation appears to be luminosity dependent (bright galaxies are dustier than faint galaxies; e.g., Bouwens et al. 2014), as well as being higher at lower redshift. So, correcting for dust would not only increase the abundance of bright galaxies more than that of faint galaxies, it would increase it by more at lower redshift, thus enhancing the differences between faint and bright galaxies at  $z > 4$ .

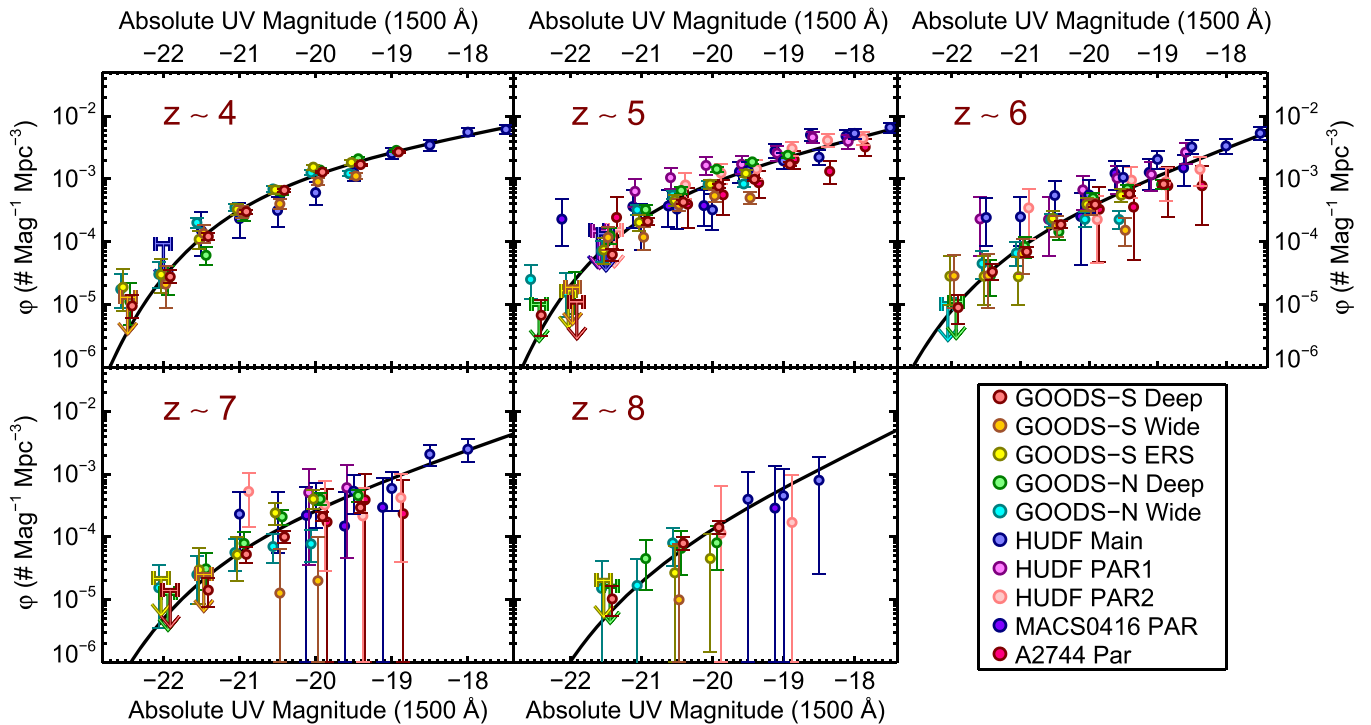
#### 6.4. Comparison to Previous Results

Our result of a similarly bright value of  $M_{UV}^*$  at  $z = 6, 7,$  and  $8$  is a dramatic change from previously published results. In

Figure 10, we show the stepwise luminosity function results from several relevant studies from the literature. Figure 15 shows our uncertainty results, highlighting both the distribution of  $M_{UV}^*$  and  $\alpha$  from the MCMC chains, as well as the covariance between the two parameters, along with previous determinations of  $M_{UV}^*$  and  $\alpha$  (Bouwens et al. 2007, 2011b; McLure et al. 2009, 2013; Ouchi et al. 2009; van der Burg et al. 2010; Schenker et al. 2013; Willott et al. 2013; Bowler et al. 2014). In this subsection we compare solely to previous work—we reserve the comparison to the contemporaneous work by Bouwens et al. (2015) to Section 6.4.1 below.

At  $z = 4$  and  $z = 5$ , both our binned luminosity function data points, as well as our Schechter function parameters, are in excellent agreement with the ground-based study of van der Burg et al. (2010). We are also in excellent agreement with the ground-based study of McLure et al. (2009) at  $z = 5$ . We find good agreement with the space-based study of Bouwens et al. (2007) at  $z = 5$ , but at  $z = 4$  the Bouwens et al. (2007) result lies outside our  $2\sigma$  confidence region on the Schechter function parameters, in that we prefer a shallower faint-end slope and a fainter value for  $M_{UV}^*$ .

At  $z = 6$ , our binned luminosity function data points are consistent within  $1-2\sigma$  with the Bouwens et al. (2007) results at the faint end (Figure 10). At the bright end, our data are *higher*



**Figure 15.** Rest-frame UV luminosity functions at each redshift for each subfield. The solid line denotes the best-fit Schechter function fit at each redshift. Upper limits are shown for magnitude bins with zero detected galaxies.

than those from both ground-based studies (although typically only different at the  $1\text{-}2\sigma$  level). This is somewhat counter-intuitive, as one may expect the ground-based studies to suffer a higher contamination rate, particularly for relatively fainter sources at higher redshift, due to their inability to resolve stars from galaxies; however, it may also be explained due to an aggressive sample selection required to minimize contamination. In any case, the differences are not highly significant, with the exception of the brightest data point from Willott et al. (2013), which gives a number density at  $M = -22.5$  of  $2.7 \times 10^{-8} \text{ Mpc}^{-3}$ . While this is consistent with our upper limit at that magnitude, it is a factor of  $\sim 250$  lower than our number density only 0.5 mag fainter at  $M = -22$  (see Table 5). Given the results at similar magnitudes at lower redshifts, it is highly unlikely that there is such a steep drop in number density over only a 0.5 mag interval, although future large area studies can better investigate the difference (Bowler et al. 2015). The larger discrepancy comes when comparing the Schechter function parameters. Specifically, Bouwens et al. (2007) found  $M^* = -20.29 \pm 0.19$ , and McLure et al. (2009) found  $M^* = -20.04 \pm 0.12$ . Both values are significantly ( $2\text{-}3\sigma$ ) fainter than our derivation of  $M^* = -21.13^{+0.25}_{-0.31}$ . For the space-based study of Bouwens et al. (2007), this is understandable, because at that time only optical data were available; thus,  $z = 6$  galaxies were selected via detections in only one band, and a robust determination of their UV absolute magnitudes was difficult. For the ground-based study of McLure et al. (2009), a cause for the difference is less clear, although the different data being used certainly plays a role.

Comparing to previous works at  $z = 7$ , we find broadly similar results, in that our results are consistent with the derived number densities from previous studies, yet our Schechter fit prefers a much brighter value of  $M_{UV}^*$ . This is easier to understand, because a number of previous studies had less data

available, and thus, utilizing smaller volumes, were unable to constrain the bright end (e.g., Bouwens et al. 2011b; Schenker et al. 2013). The exception is the recent work by McLure et al. (2013), which used a similar volume as our study, except they used the CANDELS UDS field rather than the CANDELS GOODS-N field. Examining the brightest data point from McLure et al. (2013) at  $M = -21$ , the number density is about a factor of two below our data point. However, the discrepancy is mitigated by two factors. First, as discussed by Bouwens et al. (2015), the use of fixed diameter circular apertures by McLure et al. (2013) systematically underestimates the fluxes for bright, more extended, galaxies. Bouwens et al. (2015) estimated the amplitude of this effect to be  $\sim 0.25$  mag. Shifting the brightest McLure et al. (2013) data point by 0.25 mag brings it into agreement with our results. Second, the CANDELS GOODS-N field appears to have an overdensity of  $z = 7$  galaxies. Specifically, when comparing the number density of  $z = 7$  galaxies in the GOODS-N Deep field to the GOODS-S Deep field in Figure 15, GOODS-N has a higher number density at all magnitudes. While we have not selected galaxy samples in the UDS, we can examine this further by recomputing our  $z = 7$  stepwise luminosity function using only the GOODS-S and HUDF fields. At magnitudes fainter than  $-21$ , the results do not change appreciably, as the GOODS-N Deep and Wide fields lie on either side of our Schechter fit at those magnitudes. However, the results using only GOODS-S provide a number density  $\sim 33\%$  lower at  $M = -21.5$  than our fiducial luminosity function. This difference is at the  $1\sigma$  level, due to the large Poisson noise contribution in this bin, and thus is not highly significant.

We also compare to several ground-based studies at  $z = 7$ . Ouchi et al. (2009) identified 22 bright  $z \sim 7$  candidate galaxies over  $\sim 0.4 \text{ deg}^2$ . Their data points based on detected galaxies are consistent with our own, though their strict upper limits at

$M \sim -22$  push their Schechter fit to a fainter value of  $M_{UV}^* = -20.1$ , although the large uncertainty (0.76) leaves  $M_{UV}^*$  consistent with our fit at only slightly more than the  $1\sigma$  level. The stepwise luminosity function from Castellano et al. (2010) based on deep HAWK-I data agrees well with our results, while the results from the zFourGE medium band survey of Tilvi et al. (2013) agree at  $M = -21.5$ , but differ by  $\sim 2\sigma$  at  $M = -20.5$ .

Recently, Bowler et al. (2014) made a significant improvement in search volume from the ground, discovering 34 luminous  $z \sim 7$  galaxy candidates over  $1.65 \text{ deg}^2$ , from the UltraVISTA survey data over the COSMOS field (McCracken et al. 2012) and the UKIDSS survey over the UDS field (Lawrence et al. 2007). Broadly speaking, they are consistent with our results, and are highly inconsistent with the previous determinations of  $M_{UV}^* \sim -20$  (Figure 10). There is a mild tension at  $M = -21.75$ , where the value of our Schechter fit at that point is  $2\sigma$  higher than their derived number density. However, this is their faintest magnitude bin, and is only  $\sim 50\%$  complete, thus this data point relies the most on the completeness correction. In any case, the fact that Bowler et al. (2014) found  $z \sim 7$  candidates out to very bright magnitudes gives us confidence that our brighter determination of  $M_{UV}^*$  is not necessarily dominated by cosmic variance in our fields, but is a true feature of the  $z = 7$  universe. However, this uncertainty on  $M_{UV}^*$  of  $\sim 0.4 \text{ mag}$  makes it apparent that more data are needed to constrain this parameter further.

At  $z = 8$ , we again find consistent number densities with previous studies, although our larger volume allows us to find more rare, bright ( $M = -21.5$ ) galaxies than observed in some previous surveys, pushing them to lower values of  $M_{UV}^*$  (although our uncertainty on  $M_{UV}^*$  is large, so the difference in our determination is not significantly different from previous studies). As noted, in our fit of the  $z = 8$  luminosity function we constrained  $M_{UV}^*$  to be fainter than  $-23$ , to avoid unphysically bright values, which tended to be preferred in an unconstrained fit. We note two important points when comparing to previous studies. First, while our study did not utilize the pure parallel BoRG (Trenti et al. 2011) and HIPPIES (Yan et al. 2011) programs, our bright end is consistent with that from Schmidt et al. (2014), based on a determination of the  $z = 8$  luminosity function over  $350 \text{ arcmin}^2$  of pure parallel data (for comparison, our search area at  $z = 8$  comprised  $\sim 300 \text{ arcmin}^2$ ; Table 1). The multiple sight-lines of BoRG and HIPPIES leave their results less susceptible to cosmic variance effects, so the agreement implies that cosmic variance may not be strongly affecting our bright end, although we explore this in Section 6.5.

A potentially larger difference between our results and those of previous studies is also seen at the faint end, in that our faint-end slope is possibly steeper than previously found. However, our uncertainty is large, such that our result of  $\alpha = -2.36_{-0.40}^{+0.54}$  is consistent with previous results of  $\alpha \approx -2$  (e.g., Bouwens et al. 2011b; McLure et al. 2013; Schenker et al. 2013). Previous studies use galaxies as faint as  $M = -17.5$  in their determination of the faint-end slope at  $z = 8$ . As discussed, and shown in Figure 8, we find that we fall below 50% completeness at  $M > -18.5$ , and thus we do not use galaxies fainter than that in our luminosity function determinations. While robust estimates of the number densities of galaxies at  $-18.5 \leq M \leq -17.5$  would certainly improve the confidence on the faint-end slope, we use the same deep data sets as the

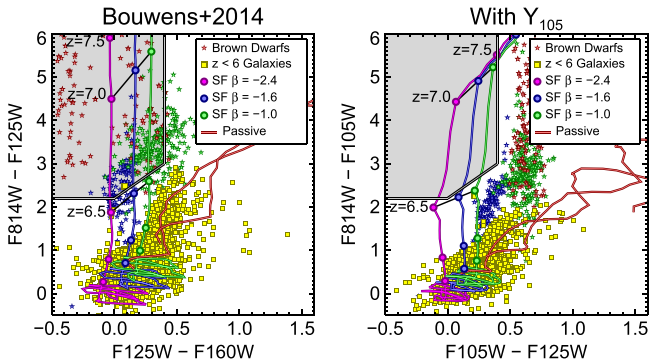
other referenced studies (HUDF). We would expect the incompleteness to be similar between all studies, although it does depend on sample selection and the exact details of the incompleteness simulations. In any case, constraints on the faint-end slope at  $z = 8$  should improve in the near future with further data from the HFF program. Our inclusion of the HFF parallel imaging, even contributing only a small number of galaxies at  $z > 7$ , did improve the fractional error on the faint-end slope ( $\sigma_\alpha/\alpha$ ) by 2% and 8% at  $z = 7$  and  $z = 8$ , respectively.

Finally, there have also been theoretical estimates of the luminosity functions at these redshifts, most prominently from Jaacks et al. (2012a), who made predictions in good agreement with our observed luminosity functions. Specifically, their simulations also predict bright values of  $M_{UV}^*$  of  $-21.15$ ,  $-20.82$ , and  $-21.00$  at  $z = 6, 7$ , and  $8$ , respectively. They also found quite steep faint-end slopes of  $-2.15_{-0.15}^{+0.24}$ ,  $-2.30_{-0.18}^{+0.28}$ , and  $-2.51_{-0.17}^{+0.27}$  at  $z = 6, 7$ , and  $8$ , respectively. Within the uncertainties, these faint-end slopes are consistent with our own, although the apparent agreement at  $z = 8$  is tantalizing (as mentioned above, we cannot constrain the slope to be so steep). Steep faint-end slopes of  $\alpha \sim -2$  at these redshifts were also seen by Salvaterra et al. (2011) and Dayal et al. (2013), but both studies also predict a brightening in  $M_{UV}^*$  toward lower redshift, which is contrary to our observations.

#### 6.4.1. Comparison to Bouwens et al. (2015)

Recently Bouwens et al. (2015) published a similar study of the evolution of the UV luminosity function at  $4 < z < 10$ . Their sample of galaxies is larger than ours—in addition to the data sets we use, they also selected galaxies from the CANDELS COSMOS, EGS, and UDS fields (although they did not use the HFF parallel fields). The data in these other CANDELS fields have a depth similar to the GOODS-S and N Wide fields, and thus are most useful for constraining the bright end of the luminosity function. Comparing our results, while the agreement at  $z = 5$  is excellent, the Bouwens et al. (2015) data points at  $z = 4$  lie at higher number densities than our own for all but the brightest bins. These differences result in a slightly steeper value of  $\alpha$  at  $z = 4$  ( $\alpha_{\text{Bouwens}} = -1.64$  versus  $\alpha_{\text{ThisStudy}} = -1.56$ ), and a slightly brighter value of  $M_{UV}^*$  ( $-20.88$  versus  $-20.73$ ).

Bouwens et al. (2015) also selected galaxy samples at  $z = 6, 7$ , and  $8$ . Broadly speaking, they found similar results to ours at  $z = 6-8$ , in that previous studies determined values of  $M_{UV}^*$  that were too faint (Figure 14). However, investigating the actual data points in Figure 10, one can see that the Bouwens et al. (2015) data points frequently lie above our own. At  $z = 7$  and  $8$ , this difference is typically significant at the  $1-2\sigma$  level, although some bins at  $z = 6$  are discrepant by up to  $4\sigma$ . At  $z = 8$ , the Bouwens et al. (2015) data points are less discrepant from our own. However, they find both a fainter value of  $M_{UV}^*$  and a shallower faint-end slope. This is primarily due to their faintest data point, which, at  $M = -17.5$ , is well below our 50% completeness limit, and lies below the extrapolation of our measured luminosity function, pushing them to a shallower slope. However, the differences at  $z = 8$  are not significant, as Figure 14 shows that the Bouwens et al. (2015) results lie comfortably within our 68% confidence contour on  $\alpha$  and  $M_{UV}^*$  (in fact, at all redshifts their Schechter parameters are consistent within the 95% confidence limits of our results). If we constrained  $\alpha$  at  $z = 8$  during our fitting to be  $> -2.3$ , we



**Figure 16.** (Left) color selection for  $z > 6.5$  galaxies used by Bouwens et al. (2015) in the COSMOS, EGS, and UDS fields, where *HST* *Y*-band data are not available. (Right) improved color selection in these fields with the addition of hypothetical *Y*-band data. Of particular note is that without *Y*-band data, the  $z > 6.5$  selection is potentially dominated by M-, L-, and T-dwarf stars. These clear out of the selection box with the addition of *Y*-band data. Additionally, galaxies with  $z < 6$  spectroscopic redshifts from the literature (yellow boxes) move farther from the selection box, and are less likely to scatter in, with the addition of the *Y*-band data.

obtain best-fit results similar to Bouwens et al. (2015; Figure 14). However, given the data at hand, there is no robust justification for such a constraint, so we do not include this in our fiducial luminosity function fits.

The three CANDELS Wide fields used only by Bouwens et al. (2015) lack space-based *Y*-band data, with *HST* data present in only four filters ( $V_{606}$ ,  $I_{814}$ ,  $J_{125}$ , and  $H_{160}$ ). These fields have deep ground-based optical data, although with much poorer angular resolution, and occasionally shallower depth than available with *HST*. Of particular worry is contamination by stars and/or brown dwarfs in these samples. The left panel of Figure 16 shows the color selection plane for galaxies at  $z > 6.5$  used by Bouwens et al. (2015) in the CANDELS COSMOS, EGS, and UDS fields. While the selection space used does include the likely colors of true  $z > 6.5$  galaxies, it also contains the bulk of M-, L-, and T-dwarf template colors. As shown in the righthand panel, by adding a single *HST* filter, the WFC3  $Y_{105}$ -band, stellar contaminants move out of the selection box, and lower-redshift galaxies move even further from the selection box. To mitigate stellar contamination, Bouwens et al. (2015) used both colors, including ground-based *Y*-band data, and the Source Extractor stellarity measurement. However, the ground-based data are presently not very deep, with Bouwens et al. (2015) typically only detecting sources with  $Y < 26$  ( $M_{UV,z=7} \leq -21$ ; see Section 3.4). Additionally, the stellarity measurement can only robustly distinguish point sources from galaxies much brighter than the detection limit. Our test with the CANDELS  $H_{160}$ -band imaging in the COSMOS and EGS fields show that a robustly identified stellar sequence in the stellarity measurement is only possible at  $H_{160} < 25$  ( $M_{UV,z=7} \leq -22$ ). In light of the apparent overabundance of bright galaxies in the Bouwens et al. (2015)  $z = 6$  and 7 samples compared to our results, we conclude that the higher quality data in our fields may yield more robust and contamination-free measurements of the number densities of bright galaxies in the distant universe.

#### 6.4.2. Previously Published Measurement Uncertainties

The differences in results, particularly on the characteristic magnitude  $M_{UV}^*$  between our current study and previous studies

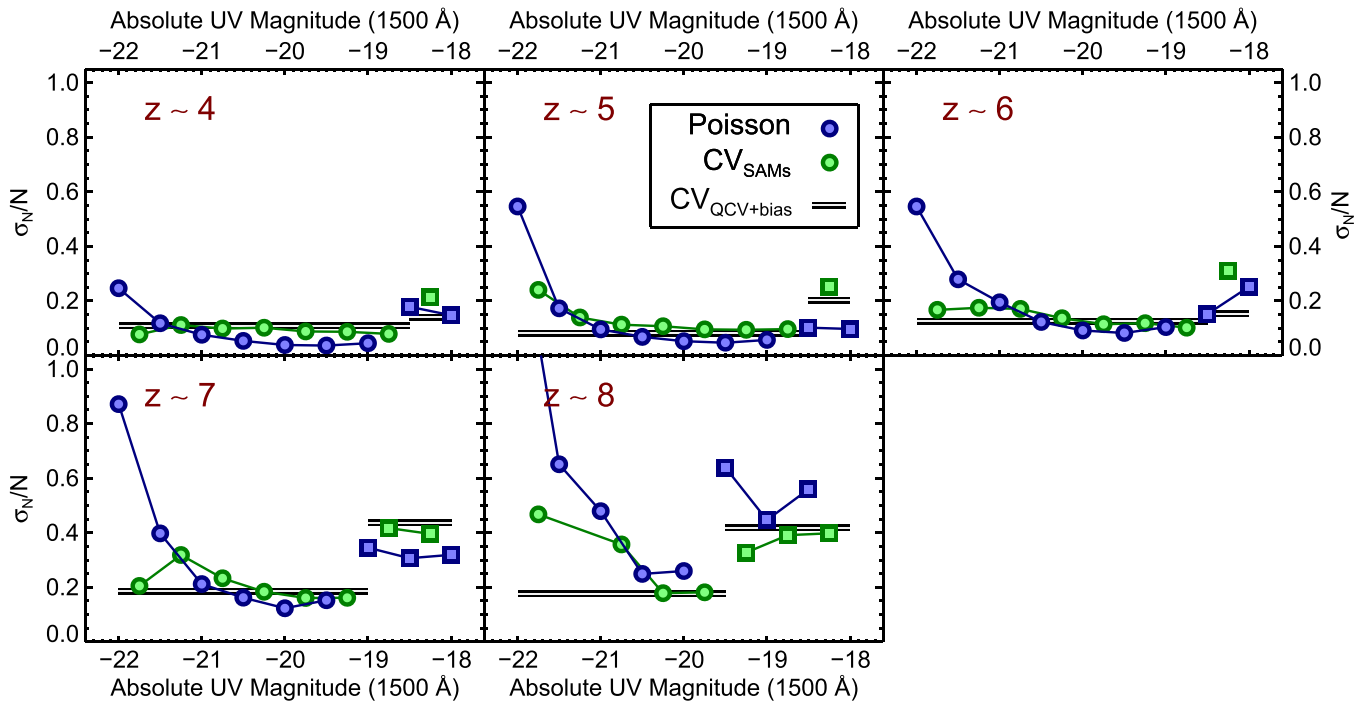
in the literature, are surprising, as in some cases the differences are larger than what would have been expected given previously published uncertainties. In particular, Bouwens et al. (2011b) initially derived  $M^* = -20.14 \pm 0.26$  at  $z = 7$ , whereas now Bouwens et al. (2015) find  $M^* = -20.87 \pm 0.26$  at  $z = 7$ , a result that is discrepant at the  $2\sigma$  level from their previous work. While the *HST* data presented in Bouwens et al. (2011b; Figure 10; green triangles) seem insufficient to constrain the bright end to such a relatively high precision (they include only the small area contained by the HUDF and ERS fields,  $< 20\%$  of the volume considered in our current work), that study made use of wide-area ground-based results when deriving their Schechter parameters to assist with constraining  $M^*$ . In particular, the data set they used to constrain the brightest magnitude bins was that of Ouchi et al. (2009). Investigating the  $z = 7$  panel of Figure 10, one can see that the two brightest Ouchi et al. data points lie systematically below not just our results, but all published results in that magnitude range (possibly due to an overestimate of the contamination; see appendix of Bouwens et al. 2015 for discussion). Thus, it may be that the inclusion of those ground-based data biased the  $M^*$  result of Bouwens et al. (2011b) to be too faint. This is confirmed by Bouwens et al. (2015), who perform a  $z = 7$  Schechter function fit using only the HUDF and ERS *HST* data, finding  $M^* = -20.6^{+0.4}_{-\infty}$ . To investigate this further, we performed a similar luminosity function fit to our data, using only data from the HUDF09 and ERS fields, finding  $M_{UV}^* = -20.64^{+0.48}_{-0.78}$  at  $z = 7$ . Thus, making use of only the pre-CANDELS *HST* data, we find a similar value for  $M_{UV}^*$  as that found by Bouwens et al. (2015) when re-examining the results from Bouwens et al. (2011b), although our uncertainty is somewhat higher. Understanding the differences in the uncertainty computations between these studies is beyond the scope of this work, but we note that our MCMC implementation was designed to produce optimal uncertainties on the Schechter function fit parameters. As shown in Figure 11, our current Schechter fit uncertainties are larger than those of Bouwens et al. (2015). While some of these differences may be due to the fact that they used a larger volume (including all five CANDELS fields), the different methods of computing the uncertainties likely play a role.

#### 6.5. Cosmic Variance

The impact of cosmic variance on our measurement of the luminosity function is minimized due to our use of several fields, which are split into four widely separated regions of the sky. However, as shown in Figure 15, there is significant variance between the different fields, particularly at  $z \geq 5$ . To estimate the effect of cosmic variance on our derived number densities, we used two techniques: a semi-empirical technique, combining the QUICKCV calculator provided by Newman & Davis (2002; using the updated version provided by Moster et al. 2011) with the recent clustering-based bias measurements from Barone-Nugent et al. (2014); and a semi-analytic technique, based on the semi-analytic models (SAMs) of Somerville et al. (2012). In Section 7 we discuss how these models, with a modification to the redshift dependence of the normalization of the dust attenuation, provide an excellent match to our measured UV luminosity functions.

For a given survey geometry, the QUICKCV program returns the fractional error in a count due to cosmic variance for an unbiased tracer of matter at a given redshift. For this





**Figure 17.** Comparison between the fractional uncertainty due to Poisson noise and that due to cosmic variance. We estimate the cosmic variance in two ways. First, we use a semi-empirical method, combining the cosmic variance estimates for an unbiased tracer of mass from QUICKCV with the clustering-based bias measurements of Barone-Nugent et al. (2014), as shown by the gray bars (for bright and faint galaxies). Second, we estimate cosmic variance uncertainties by examining the variation in the number of galaxies as a function of rest-frame absolute UV magnitude from a set of semi-analytic models (discussed in Section 7), with the volume approximated as that of the two CANDELS/GOODS fields, except at faint magnitudes, where we use a HUDF-sized volume. The Poisson values shown come from the uncertainties on the number densities shown in Figure 10. The circles and squares denote magnitudes where the majority of our galaxies come from the CANDELS fields and HUDF field, respectively. The two estimates of the cosmic variance uncertainty show very good agreement. In nearly all cases, the cosmic variance uncertainty is greater than the Poisson uncertainty, thus cosmic variance is likely not the dominant source of uncertainty in our measured luminosity functions.

calculation, we estimated the fractional error separately for GOODS-S, GOODS-N, MACS-0416 parallel, and Abell 2744 parallel fields, adding the variances in quadrature to derive a final value of  $\sigma_{CV}$  for a given redshift bin. In the GOODS-S field, we included the area from the three HUDF09 fields, because even the parallel fields are separated by only a few arcmin from the GOODS-S proper. For the input survey geometries, we estimate rectangular regions of the approximate shape of the GOODS fields, with an enclosed area equal to the GOODS-S Deep+Wide+ERS+HUDF09 fields for GOODS-S, and the GOODS-N Deep+Wide for GOODS-N. The field geometries were thus  $10'.2 \times 15'.03$  for GOODS-S,  $9'.51 \times 14'.65$  for GOODS-N, and  $2'.1 \times 2'.1$  for each of the HFF parallel fields. However, at faint magnitudes, our galaxy sample primarily comes from the HUDF, thus we also estimated the QUICKCV-derived cosmic variance uncertainty with the HUDF area only, with a geometry of  $2'.26 \times 2'.26$ . To convert these unbiased estimates of cosmic variance to values appropriate to our galaxy sample, we use the recently published clustering-based bias measurements from Barone-Nugent et al. (2014), who used the galaxy sample of Bouwens et al. (2015) for their calculation (because they did not measure the clustering at  $z = 8$ , we use the  $z = 7$  bias values for our  $z = 8$  cosmic variance estimate). They estimated the bias for both bright and faint galaxies, splitting their sample at  $M_{UV} = -19.4$  at each redshift. Our estimates of the fractional uncertainty on galaxy counts due to cosmic variance from this method are shown as the gray bars in Figure 17, where we show values of this quantity for both bright and faint galaxies.

For our semi-analytic cosmic variance estimate, we used mock catalogs of the Somerville et al. (2012) SAMs, which cover an area  $\sim 40\times$  larger than that of the combined CANDELS/GOODS fields. We extract independent, GOODS-sided volumes from these catalogs, exploring the variation in number counts in the independent volumes as a function of UV absolute magnitude. At magnitudes brighter than  $-18.5$  at  $z = 4-6$  ( $-19$  at  $z = 7$ ;  $-19.5$  at  $z = 8$ ) we estimated our survey as being two  $10' \times 16'$  fields, representing a combination of the CANDELS/GOODS fields with the five single-WFC3 pointing fields (HUDF09 and HFF). At fainter magnitudes, where the majority of our objects come from the HUDF, we assume a single  $2'.26 \times 2'.26$  field. We calculate the  $1\sigma$  fractional uncertainty on the number density,  $\sigma_{cv}/N$ , by bootstrap resampling galaxies in a given  $M_{UV}$  bin at each redshift. This 1-sigma fractional uncertainty includes the Poisson noise, thus we subtract the Poisson errors using the recipe of Gehrels (1986) to calculate the fractional uncertainty on the number density due to cosmic variance only. The uncertainty for the total survey volume is calculated by adding the variance for two GOODS-sized fields in quadrature for bright bins (and the HUDF-only for faint bins), and is shown in Figure 17.

Comparing the SAM-derived cosmic variance values to those from QUICKCV, we find generally excellent agreement. The SAM method predicts, as expected, a larger cosmic variance uncertainty for the brightest galaxies, although this is understood as the Barone-Nugent et al. (2014) “bright” sample encompassed galaxies down to  $M_{UV} = -19.4$ , and thus likely has a relatively faint median magnitude. Future measurements

of the bias in finer-resolution bins of bright galaxies can probe this effect further. The agreement at the faint end is generally good as well, with the exception of  $z = 6$ , where the semi-empirical method predicts a somewhat low uncertainty due to the Barone-Nugent et al. (2014) bias measure at  $z = 6$ , which is slightly lower than at  $z = 5$ . While both of these methods are estimates, and thus may not be extremely accurate, the general agreement between these two different techniques implies that our estimates of the cosmic variance uncertainty are not highly inaccurate.

To assess the impact of cosmic variance on our measured luminosity functions, we compared both of our estimates of the cosmic variance uncertainty to the Poisson noise from our stepwise luminosity functions, as shown in Figure 17. At all redshifts, the data at  $M < -21$  are dominated by Poisson noise ( $M < -20.5$  at  $z \geq 6$ ), thus we do not expect cosmic variance to be dominating the uncertainties on the bright end of the luminosity functions derived here. However, cosmic variance may play some role at the faint end, where we are restricted to small fields. At  $z = 7$ , there does appear to be a step in the stepwise luminosity function at  $M \geq -18.5$ , where the  $M = -19$  point is below our best-fit Schechter function, and the  $M = -18.5$  point is above. At  $M \geq -18.5$  our data come from only the HUDF main field, thus this break represents the point where we become reliant on a single small field. As shown in Figure 17, the cosmic variance uncertainty at faint magnitudes at  $z = 7$  is  $\sim 40\%$ , which is somewhat larger than the Poisson uncertainty; thus, cosmic variance may bear some responsibility for this discontinuity in the luminosity function at the faint end. Future measures of the luminosity function at  $M \geq -18.5$  from the HFF lensing program may improve these constraints, but while faint galaxies may be found, the volumes will still be incredibly small (e.g., Robertson et al. 2014). Thus, robust constraints on the number densities at this faint level at  $z \geq 7$  may need to wait until the *JWST*. We conclude that while the effects of cosmic variance are not negligible, they are not the dominant source of uncertainty on the abundances of the bright objects we have discovered at very high redshift.

### 6.6. Do the Data Support a Schechter Function?

When allowed to choose any value of  $M_{UV}^*$ , our  $z = 8$  Schechter function fit preferred very bright values of  $M_{UV}^*$ , such that all observed data points lay on the faint-end slope part of the function. This implies that the  $z = 8$  luminosity function is consistent with a single power law. Such a functional form is what one might expect when the feedback effects that govern the bright end at lower redshift (i.e., mainly feedback due to accreting supermassive black holes) disappear, or if dust attenuation ceases to be a factor. Bowler et al. (2014) recently postulated that the  $z = 7$  luminosity function is better fit by a double power law, rather than a Schechter form. At  $z = 7$ , our stepwise data appear consistent with the Schechter fit out to the brightest magnitudes we cover. To see whether our data show a preference for a Schechter functional form at all redshifts, we performed three fits to the data: a Schechter fit, a single power law, and a double power law. To place these fits on equal ground, we found the best-fit parameters for each function using a simple maximum likelihood routine. For the Schechter fit, we used the function shown in Equation (4), investigating a range of  $M_{UV}^*$  with  $\Delta M = 0.1$  mag, and  $\alpha$  with  $\Delta\alpha = 0.02$ . We approximated a single power law using the Schechter functional form with  $M_{UV}^*$  fixed at  $-30$ . For the double power

law, we used the form given in Equation (2) of Bowler et al. (2014), which is similar to the Schechter function at the faint end, but replaces the bright end with a second power law with slope  $\beta$ . In all cases,  $\varphi^*$  is found as the normalization, such that the total number of expected objects for a given function is equal to the number observed. The likelihood that a given functional form represents our data was calculated in an identical manner as in Section 5.1, using Equations (5) and (6).

To compare the results from these fits at each redshift, we used the Bayesian information criterion (BIC). This is similar to a  $\chi^2$  statistic, except that it takes into account both the number of data points and the number of free parameters. For a model to be preferred over a competing model, it must have a BIC lower by at least 2. This is sensible, because adding a free parameter must yield a better fit for that model to be preferred. The BIC is calculated as

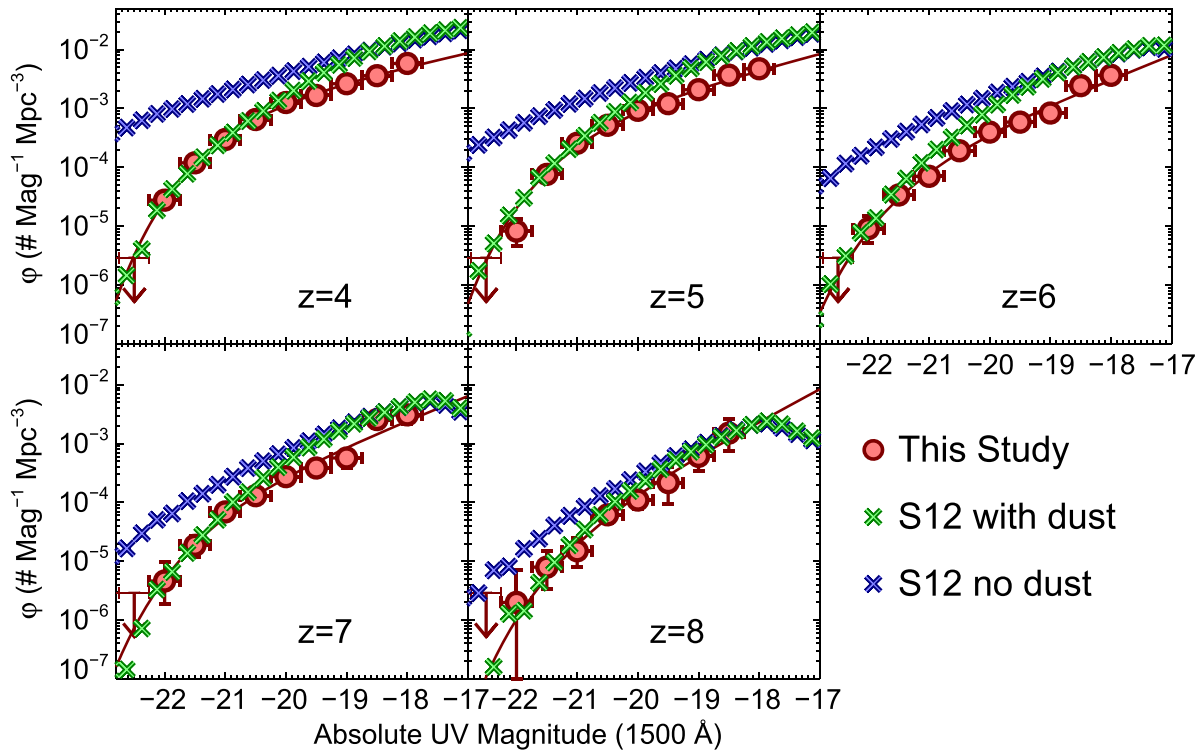
$$\text{BIC} = -2 \ln(\mathcal{L}) + k \ln(N) \quad (8)$$

where  $N$  is the number of data points and  $k$  is the number of free parameters (Liddle 2004). For the Schechter, double power law, and single power law fits, the number of free parameters are 2 ( $M_{UV}^*$ ,  $\alpha$ ), 3 ( $M_{UV}^*$ ,  $\alpha$ ,  $\beta$ ), and 1 ( $\alpha$ ), respectively (we do not count  $\varphi^*$  as a free parameter because it is a normalization). The number of data points is the number of galaxies from our sample used in the fit, which is restricted to those brighter than the 50% completeness limits discussed previously. This gives  $N = 2788, 1812, 605, 221$ , and 47 galaxies at  $z = 4, 5, 6, 7$ , and 8, respectively.

The results from this analysis are shown in Table 6. A difference in the absolute value of the BIC of 2 is interpreted as “positive” evidence, whereas a difference of 6 or higher is “strong” evidence, both in favor of the model with the smaller value. In Table 6, in addition to the value of the BIC, we show the difference between the BIC values for the Schechter versus double power law, and Schechter versus single power law. In this formalism, a negative difference is in favor of the Schechter function. Comparing the Schechter function versus the double power law, we find that a Schechter form is strongly preferred to either a double or single power law at  $z = 4-7$ . This is not surprising; there is clearly a deficit of observed galaxies at the bright end when compared to the best-fit power law (Figure 10). However, no such deficit is visible at  $z = 8$ , and this is confirmed, as both the Schechter fit and the single power law fit have effectively identical values as the BIC. We conclude that our data support an exponential decline at the bright end of the luminosity function at  $z = 4, 5, 6$ , and 7. At  $z = 8$ , we do not see any evidence for a decline in the bright end, at least out to  $M = -21.5$ . Further data are needed to show whether one can either detect, or rule out, a decline at the bright end at  $z = 8$ . If the latter is true, it could indicate a significant change in the halo masses of bright galaxies, a drop in dust attenuation in bright galaxies, or a change in the physics governing the feedback in bright galaxies in the distant universe.

## 7. COMPARISON TO SAM PREDICTIONS

In this section we compare our observations with predictions from theoretical models set within the predominant  $\Lambda$  Cold Dark Matter ( $\Lambda$ CDM) paradigm. All such models, whether based on numerical hydrodynamics or semi-analytic techniques, currently rely upon phenomenological “subgrid” recipes



**Figure 18.** Rest-frame ultraviolet luminosity functions at  $z = 4-8$ , comparing our observations to the semi-analytic models (SAMs) of Somerville et al. (2012; S12). The crosses mark the models of S12 both with and without dust. In order for the models to be consistent with the observations, it is clear that some dust is needed at all redshifts (except perhaps at  $z = 8$ ), particularly at the bright end. Using a simple model for dust attenuation, our results suggest that the dust-to-metal ratio or dust geometry must change as a function of redshift, which is a continuation of a trend already pointed out in previous studies. The turnover at very faint luminosities in the SAM predictions is due to the resolution limit of the *Bolshoi* simulation.

**Table 6**  
Comparison of Luminosity Function Fits

Redshift	BIC Schechter	BIC Double	BIC Power	$\Delta$ BIC Sch-Dou	$\Delta$ BIC Sch-Pow
4	358	377	641	-18.5	-283
5	540	562	694	-21.7	-153
6	350	361	376	-11.1	-25.7
7	225	234	235	-9.28	-9.83
8	86.9	92.3	86.7	-5.36	-0.26

**Note.** The comparison of the Bayesian information criterion statistic for fits to our  $z = 4, 5, 6, 7$ , and  $8$  luminosity functions using a Schechter, double power law, and single power law functional form. A difference in the absolute value of BIC between two models of  $\geq 2$  (6) is positive (strong) evidence for the preference of one model over another. A Schechter function is strongly preferred over a single power law at all redshifts except  $z = 8$ , where our data cannot distinguish between the two models.

to treat the physics on scales smaller than those that can be directly resolved. These processes include star formation and feedback from massive stars, supernovae, and supermassive black holes. The phenomenological recipes are parameterized and must be empirically calibrated. Here, we compare our new observations at  $z = 4-8$  with predictions from the models presented by Somerville et al. (2012; hereafter S12). The subgrid recipes in these models have been calibrated using a set of observations at  $z \sim 0$ , and Somerville et al. (2012) presented a comparison with available observations from  $z \sim 0-5$ . It is therefore very interesting to test these model predictions—with no re-tuning of the free parameters

controlling physical processes—in the higher redshift regime probed in this work.

Figure 18 shows our estimates of the rest-UV luminosity function compared with the S12 SAM predictions with and without dust. It is already interesting that the dust-free model predictions are even in plausible agreement with the observations (i.e., the model predictions lie above the observations at all luminosities and redshifts). Next, we can ask which characteristics the dust extinction must have in order to be consistent with the observations. One can immediately see that the dust extinction must be differential with both luminosity (more luminous galaxies are more extinguished) and redshift (galaxies are less dusty at higher redshift). We use a simple approach to model the dust extinction: as in S12, we assume that the face-on dust optical depth in the  $V$ -band is given by  $\tau_{0,V} = \tau_{\text{dust}} m_{\text{cold}} Z_{\text{cold}} / r_{\text{gas}}^2$ , where  $m_{\text{cold}}$  is the mass of cold gas in the disk,  $Z_{\text{cold}}$  is the metallicity of the cold gas,  $r_{\text{gas}}$  is the exponential scale radius of the gaseous disk, and  $\tau_{\text{dust}}$  is a normalization parameter. The values of  $m_{\text{cold}}$ ,  $Z_{\text{cold}}$ , and  $r_{\text{gas}}$  are predicted by the SAM. We treat  $\tau_{\text{dust}}$  as a free parameter. We then assign random inclinations to our galaxies and use a “slab” model to compute the inclination-dependent extinction (see S12 for details). We used a Calzetti attenuation curve to compute the attenuation at  $1500 \text{ \AA}$  at each redshift.

In S12, we showed that if we normalize  $\tau_{\text{dust}}$  to match observations at  $z \sim 0$  and use a fixed value, our model overpredicts the dust extinction at higher redshift. Similarly, we find here that the empirical redshift-dependent function for  $\tau_{\text{dust}}$  adopted in S12 based on observations at  $z \lesssim 5$  overpredicts the extinction at  $z \gtrsim 5$ . When we empirically adjust  $\tau_{\text{dust}}$  to obtain a good fit to the bright end of the observed LF in the five

redshift bins shown, we find that  $\tau_{\text{dust}} \propto \exp(-z/2)$  produces a reasonably good fit over this redshift range, where  $z$  is redshift. This may be physically interpreted as either a changing dust-to-metal ratio, or a systematic evolution in the dust geometry relative to our simple slab model. The required luminosity and redshift dependence of the dust extinction is in qualitative agreement with observational conclusions drawn based on the UV colors (Finkelstein et al. 2012b; Bouwens et al. 2014). While the agreement at the bright end is excellent, the predicted faint end is too steep, particularly at lower redshift. This could imply that the predicted luminosity dependence of the dust attenuation results in too little attenuation at faint magnitudes, or it could reflect on the impact of feedback on the star formation in the simulations. Performing a similar analysis with stellar mass functions and UV luminosity functions in tandem can break this degeneracy.

In future works, we plan to investigate whether the dust extinction parameters derived from SED fitting on the observations are consistent with the empirical SAM requirements. In addition, we plan to use these models, which plausibly match the observed UV luminosity functions, to make predictions for the clustering, stellar fractions, and other properties of high-redshift galaxies. We will also show the results of varying the subgrid recipes for star formation and feedback, to illustrate what physical insights can be gained from these observations. For the moment, however, it is intriguing that the models developed to explain galaxies at a very different epoch are plausibly consistent with these new observations.

## 8. EVOLUTION OF THE COSMIC SFR DENSITY

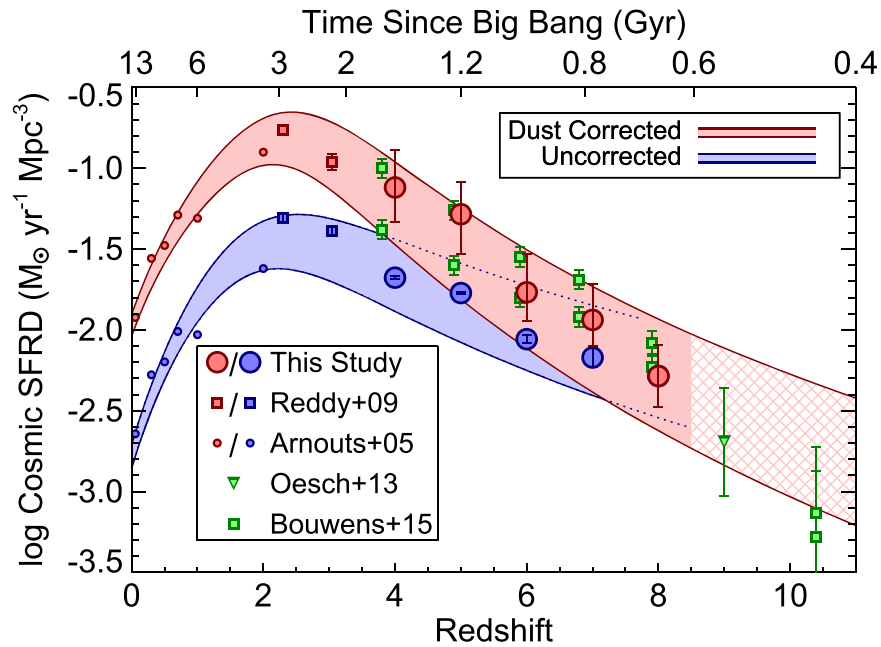
While the evolution of the shape of the luminosity function can provide interesting constraints on the physics of galaxy evolution, the integral of the luminosity function provides a key measure of the total number of UV photons produced at a given redshift. This is a key constraint in two ways. First, the integral of the total SFR density provides a key check against the measured stellar mass density (e.g., Bouwens et al. 2014). Second, assuming a conversion between UV and ionizing photons, this measure can determine whether galaxies are producing enough ionizing photons to reionize the universe at a given redshift (e.g., Madau et al. 1996; Finkelstein et al. 2012a; Robertson et al. 2013).

We calculated the luminosity density at each redshift, integrating down to  $M_{\text{UV}} = -17$ . This is approximately the magnitude of the faintest galaxy in our  $z = 8$  sample, and also facilitates comparison with recent works that use a similar magnitude limit. Galaxies likely exist beyond this magnitude limit (e.g., Trenti et al. 2012; Alavi et al. 2014), which we will consider in the next subsection. We utilized the results of our MCMC luminosity function fitting chain to derive a robust estimate of both the rest-frame UV specific luminosity density ( $\rho_{\text{UV}}$ , in units of  $\text{erg s}^{-1} \text{Hz}^{-1} \text{Mpc}^{-3}$ ) and its uncertainties. In each step of the chain, we calculated  $\rho_{\text{UV}}$  by taking the luminosity function from the best-fit Schechter function parameters for that step, and integrating it from  $-23 < M_{1500} < -17$ . To convert this number to an SFR density, we use the relation adapted from Kennicutt (1998,  $\rho_{\text{SFR}} = 1.25 \times 10^{-28} \rho_{\text{UV}}$ ), which converts the specific UV luminosity density to a SFR density ( $\rho_{\text{SFR}}$ ), assuming a Salpeter IMF and a constant SFH over  $\geq 100$  Myr. The original coefficient from Kennicutt (1998) was 1.4; however, updated

stellar population models (e.g., Bruzual & Charlot 2003) imply a somewhat smaller value. We chose a value of 1.25 to be consistent with the assumptions used in Bouwens et al. (2015), although we note that an even lower coefficient of 1.15 was used by Madau & Dickinson (2014). The quoted value of  $\rho_{\text{UV}}$  or  $\rho_{\text{SFR}}$  is the median of the values recorded from all of the MCMC steps, while the 68% confidence range is taken to be the central 68% of values.

Although the UV luminosity is a relatively easy observable in this epoch, the major drawback in its use as an SFR indicator is its susceptibility to attenuation by dust. As a bevy of recent work has shown, this dust correction is important even out to  $z \sim 7-8$  (e.g., Finkelstein et al. 2012b; Dunlop et al. 2013; Bouwens et al. 2014, see also Section 7). To calculate the total SFR density, we corrected the observed SFR density for extinction using a new iteration of SED fitting, including the deep *Spitzer*/IRAC data (Section 3.6), which is a crucial probe of the rest-frame optical light, providing better constraints on the dust attenuation. Using these updated extinction results, we calculated a sigma-clipped median and standard deviation for the best-fit extinction values at a given redshift in four magnitude bins:  $< -21$ ,  $-21$  to  $-20$ ,  $-20$  to  $-19$ , and  $-19$  to  $-17$ . We recover previously observed trends that dust extinction lessens with both increasing redshift and decreasing UV luminosity (e.g., Finkelstein et al. 2012b; Bouwens et al. 2014). The values of  $E(B - V)$  for bright galaxies decreases from 0.15 at  $z = 4$  to 0.02 at  $z = 7$ , and for faint galaxies from 0.06 at  $z = 4$  to 0.0 at  $z = 7$ . The small numbers, limited wavelength coverage, and faint magnitudes of  $z = 8$  galaxies make it difficult to measure their extinction, therefore we assumed  $E(B - V) = 0$  for all  $z = 8$  galaxies. The spread in  $E(B - V)$  values at all redshifts and luminosities is  $\sim 0.1$ , thus we assume this value in all cases (with the exception of  $z = 8$ , where we fix  $E(B - V)$  to zero). To include the uncertainty in  $E(B - V)$  in our derived dust-corrected SFR density, in each step of the chain we draw a new value of  $E(B - V)$  for a given redshift and magnitude bin, modifying the fiducial value by a number drawn from a Gaussian distribution with a standard deviation equal to the  $E(B - V)$  spread of 0.1. The values of  $\rho_{\text{UV}}$  and the observed and dust-corrected values of  $\rho_{\text{SFR}}$  are given in Table 7. These values do not include potential submillimeter galaxies, which lie below our rest-frame UV detection limits. However, because we are observing at  $z > 4$ , we expect their impact on the total cosmic SFR density to be minimal (see Table 7 from Bouwens et al. 2012).

In Figure 19, we show our derived values of the cosmic SFR density. Our results are, for the most part, consistent with those of Bouwens et al. (2015), although our observed values are lower by a few  $\sigma$  at  $z = 4$  and 5, likely due to our shallower faint-end slopes at these redshifts. To study the evolution of  $\rho_{\text{UV}}$  with redshift, we fit the function provided by Madau & Dickinson (2014), which is given in Equation (7). Although we have included lower-redshift data in our fit, we do not discuss here our results for the low-redshift slope or the peak redshift; these are better obtained from Madau & Dickinson (2014), who use a compilation of several sources, including far-infrared observations. As we are adding data at high redshift, it is interesting to examine the trends there. We find that at  $z > 3$  the uncorrected values of  $\rho_{\text{UV}}$  evolve as  $(1 + z)^{-2.4 \pm 0.3}$ , while the dust-corrected values evolve as  $(1 + z)^{-4.3 \pm 0.5}$ . Most interesting, the observed trend of the evolution of the total



**Figure 19.** Evolution of the cosmic star-formation rate density, derived by integrating the best-fit Schechter function at all redshifts to  $M_{UV} < -17$ . Our data are shown as large circles. To extend this analysis to lower redshifts, we also show the values at  $z \sim 2-3$  from Reddy & Steidel (2009), and from  $z = 0-2$  from Arnouts et al. (2005). For both studies, we integrated the published best-fit Schechter function parameters to  $-17$  to derive the uncorrected values of  $\rho_{UV}$ . We used the published ratio of the dust-corrected-to-unobserved values of  $\rho_{UV}$  from Reddy & Steidel (2009) to calculate the dust-corrected values at  $z \sim 2-3$ . At  $z \leq 2$ , we used the dust correction from Schiminovich et al. (2005), which assumes a constant value of  $A_{UV} = 1.8$  at all redshifts. We used Equation (7) to fit the observed trends, deriving the uncertainties on this fit via  $10^4$  Monte Carlo simulations, shown as the shaded region the 68% confidence range from these fits. The total (dust-corrected) SFR density evolves with  $(1+z)^{-4.3 \pm 0.5}$  from  $z = 3-8$ . The green symbols show the high-redshift results from Bouwens et al. (2015) and Oesch et al. (2013), which were not included in the fit, but are consistent with the observed trends at the  $\sim 1\sigma$  level.

**Table 7**  
Rest-frame UV Luminosity Densities and SFR Densities

Redshift	$\log \rho_{UV}$	$\log$ SFR Density	
	( $\text{ergs s}^{-1} \text{Hz}^{-1} \text{Mpc}^{-3}$ )	( $M_{\odot} \text{yr}^{-1} \text{Mpc}^{-3}$ )	( $M_{\odot} \text{yr}^{-1} \text{Mpc}^{-3}$ )
	Observed	Observed	Dust-corrected
4	$26.26^{+0.01}_{-0.01}$	$-1.59^{+0.01}_{-0.01}$	$-1.03^{+0.23}_{-0.21}$
5	$26.17^{+0.01}_{-0.01}$	$-1.69^{+0.01}_{-0.01}$	$-1.20^{+0.20}_{-0.25}$
6	$25.88^{+0.02}_{-0.02}$	$-1.97^{+0.02}_{-0.02}$	$-1.68^{+0.24}_{-0.18}$
7	$25.77^{+0.06}_{-0.06}$	$-2.09^{+0.06}_{-0.06}$	$-1.85^{+0.22}_{-0.16}$
8	$25.65^{+0.19}_{-0.19}$	$-2.20^{+0.19}_{-0.19}$	$-2.20^{+0.19}_{-0.19}$

**Note.** All values have been computed down to  $M_{UV} = -17$ . The dust correction was derived based on the values of  $E(B - V)$  derived from SED fitting, with the dust-corrected SFR densities including an uncertainty term from the spread of extinction values at a given absolute magnitude. The SFRs were computed assuming the Kennicutt (1998) conversion from the UV luminosity density ( $\rho_{UV}$ ), assuming a Salpeter IMF, and a constant star-forming population with age  $\geq 100$  Myr.

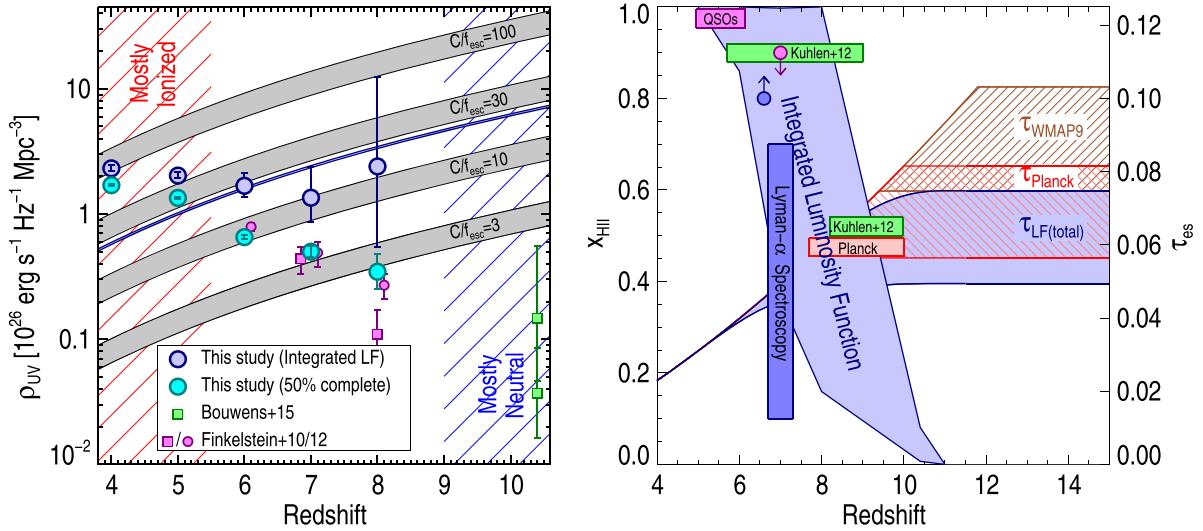
SFR density is consistent within  $1\sigma$  of the published results at  $z = 9$  from Oesch et al. (2013) and at  $z = 10$  from Bouwens et al. (2015). Thus, the Oesch et al. (2013) and Bouwens et al. (2015) results are consistent with a smooth extrapolation of our derived cosmic SFR density at  $4 < z < 8$  to higher redshift, with no break.

### 8.1. Constraints on Reionization

Although it is presently generally assumed that galaxies dominated the ionizing photon budget for the reionization of the IGM, it has been difficult for observations to obtain robust proof. Analyses of the IGM via line-of-sight quasar

observations have been able to show that reionization was likely complete by  $z \sim 6$  (e.g., Fan et al. 2006; although see Mesinger 2010 and Becker et al. 2015). Additionally, observations of the cosmic microwave background (CMB) radiation constrain the total optical depth due to electron scattering, which, while it cannot directly inform the duration of reionization, can give an estimate of the reionization redshift ( $z_{\text{reion}}$ ) if reionization was instantaneous. The results from the *WMAP* nine-year data set give  $\tau_{\text{es}} = 0.088 \pm 0.014$ , which corresponds to  $z_{\text{reion}} = 10.6 \pm 1.2$  (Hinshaw et al. 2013), while the recent *Planck* results suggest  $z_{\text{reion}} = 8.8^{+1.2}_{-1.1}$  (Planck Collaboration et al. 2015).

The primary reason for the current uncertainties in the contribution of galaxies to reionization lies in the uncertainty in the faint-end slope measurements, as well as in the assumptions of the escape fraction of ionizing photons ( $f_{\text{esc}}$ ) and the clumping factor in the IGM ( $C$ ). The clumping factor is primarily constrained theoretically, but most studies agree that it is low ( $< 6$ ) at high redshift (e.g., Faucher-Giguère et al. 2008; Pawlik et al. 2009; McQuinn et al. 2011; Finlator et al. 2012). To infer a number of escaping ionizing photons from observations of galaxies, one first needs to take the observed UV light and assume an IMF and a metallicity. Then, to calculate the number of these ionizing photons available for reionization, multiply by an assumed value of  $f_{\text{esc}}$ . However, it is difficult to constrain  $f_{\text{esc}}$  directly at high redshift, because the correction for intervening IGM absorption systems is extremely high at  $z > 4$ . Significant effort is being expended on observationally constraining  $f_{\text{esc}}$  at  $z < 4$ . Although bright galaxies at  $z \sim 1$  have very low escape fractions (relative to the UV emission) of  $f_{\text{esc,rel}} < 2\%$  (Siana et al. 2010), escaping ionizing emission has been observed from small fractions of



**Figure 20.** (Left) specific luminosity density ( $\rho_{UV}$ ) vs. redshift (similar to Figure 3 from Finkelstein et al. 2012a). Here we show our luminosity functions integrated down to  $M < -13$  as blue circles. The cyan circles denote the value of  $\rho_{UV}$  when we integrate down to our 50% completeness limit ( $-18$  at  $z = 7$ ). Recent results from Bouwens et al. (2015) at  $z \approx 10$  are shown in green, with the lower and upper squares representing limiting magnitudes of  $-17$  and  $-13$ , respectively. The wide gray curves denote the value of  $\rho_{UV}$  needed to sustain a fully reionized IGM at a given redshift, for a given ratio of the clumping factor  $C$  over the escape fraction of ionizing photons  $f_{esc}$  (Madau et al. 1999). The thin blue curve shows our fiducial value of  $C = 3$  and  $f_{esc} = 13\%$ . (Right) The volume ionized fraction,  $x_{HII}$ , of the IGM, which can be sustained given the integral of our luminosity functions at  $z = 4-8$  (as well as that at  $z = 10.4$  from Bouwens et al. 2015). We assume that the luminosity function extends to  $M_{UV} = -13$ ,  $C = 3$ , and  $f_{esc} = 13\%$  (this escape fraction is the highest that does not violate constraints set by the Ly $\alpha$  forest at  $z = 6$ ; Finkelstein et al. 2012b). We plot constraints on  $x_{HII}$  from the spectroscopy of quasars at  $z < 6$  from Fan et al. (2006) and at  $z = 7$  from Bolton et al. (2011). The blue circle denotes constraints on  $x_{HII}$  from the evolution in the Ly $\alpha$  luminosity function from  $z = 5.7$  to  $6.6$  from Ouchi et al. (2010), while the blue bar denotes the range of  $x_{HII}$  values inferred from  $z \approx 7$  follow up Ly $\alpha$  spectroscopic studies (e.g., Faisst et al. 2014; Pentericci et al. 2014; Tilvi et al. 2014). The instantaneous redshift for reionization from *Planck* ( $8.8^{+1.2}_{-1.1}$ ) is indicated by the red rectangle. The derived 50% and 90%  $x_{HII}$  redshifts from the study of Kuhlen & Faucher-Giguère (2012) are shown in green. The righthand axis corresponds to the hatched regions, which show the Thomson optical depth to electron scattering ( $\tau_{es}$ ), as predicted by our integrated luminosity functions (blue) compared to *Planck* (red) and WMAP9 (brown). Compared to previous results, the improved constraints on the luminosity functions yield a tighter range of possible reionization histories. Broadly speaking, we find a picture where the universe is fully ionized by  $z = 6$ , with the neutral fraction becoming non-negligible at  $z \geq 7$ , with  $\tau_{es}$  highly consistent with the *Planck* value.

galaxies probed by studies at  $z \sim 3-4$  (e.g., Steidel et al. 2001; Shapley et al. 2006; Iwata et al. 2009; Vanzella et al. 2010; Nestor et al. 2011); though some ground-based studies may suffer from contamination by intervening sources (e.g., Vanzella et al. 2012). Recent results imply that escape fractions from star-forming galaxies at  $z \sim 2-3$  range from 5% to 20%, with lower-mass galaxies, especially those with Ly $\alpha$  in emission, having a greater likelihood of having detectable escaping ionizing emission (e.g., Mostardi et al. 2013; Nestor et al. 2013).

Finkelstein et al. (2012a) used measurements of the emission rate of ionizing photons from observations of the Ly $\alpha$  forest in quasar spectra to place an upper limit on  $f_{esc}$  from galaxies. Assuming that the rest-frame UV luminosity function extended down to  $M_{UV} = -13$ , the escape fraction must be  $f_{esc} < 13\%$  to avoid violating the Ly $\alpha$  forest measurements of Bolton & Haehnelt (2007). Using this value, and assuming  $C = 3$ , the luminosity functions available at the time were consistent with a wide range of reionization histories, including an end redshift as late as  $z \lesssim 5$ , and an ionized fraction at  $z \sim 7$  from 30% to 100%. Kuhlen & Faucher-Giguère (2012) and Robertson et al. (2013) did similar analyses, folding in additional observables (e.g., the Ly $\alpha$  forest and CMB), and found that in order to complete reionization by  $z \sim 6$ , the luminosity function must extend much deeper than can presently be observed, and/or the average escape fraction must be higher at higher redshift.

Here, we use our updated luminosity functions to re-examine the contribution of galaxies to reionization. Figure 20 shows both the observable specific UV luminosity density ( $\rho_{UV}$ ), which we define to be that above our 50% completeness limit,

as well as the total  $\rho_{UV}$ , which we define as the integrated luminosity function down to  $M_{1500} = -13$ . We then compare these values to the critical number of UV photons necessary to sustain an ionized IGM at a given redshift, taken from Madau et al. (1999). This figure is similar to Figure 3 from Finkelstein et al. (2012a), thus we refer the reader there for more details. Effectively, these critical curves depend on assumptions about the stellar IMF, metallicity,  $f_{esc}$ , and clumping factor. The first two are responsible for the conversion from observed UV photons to intrinsic ionizing photons. We assumed a Salpeter IMF, and the width of the curves denote the impact of changing the metallicity from  $0.2 \leq Z/Z_{\odot} \leq 1.0$ . We show several curves for the reader's choice of the ratio of  $C/f_{esc}$ . Here, we use a fiducial value of  $C = 3$  and  $f_{esc} = 13\%$ , which are consistent with Finkelstein et al. (2012a).

The right panel of Figure 20 shows the ionization history of the IGM, comparing our derived value for the total specific UV luminosity density to our fiducial model of  $C = 3$  and  $f_{esc} = 13\%$ , folding in the values at  $z = 10.4$  from Bouwens et al. (2015) to extend our analysis beyond  $z = 8$ . Our luminosity functions are consistent with a reionization history that starts at  $z \sim 11$ , and ends by  $z > 5$ . Although the exact value of the volume ionized fraction in the IGM is uncertain between these redshifts, due to the persistent uncertainty in the faint-end slope, our results imply the following constraints (given the caveat of our assumptions). At  $z = 6$ , we can constrain  $x_{HII} > 0.85$  ( $1\sigma$ ), while out to the limit of our observations at  $z = 8$  the data are still consistent with a fully ionized IGM (68% C.L. of  $0.15 < x_{HII} < 1.0$ ). We find a midpoint of reionization ( $x_{HII} = 0.5$ ) of  $6.7 < z < 9.4$  (68% C.L.).

Broadly speaking, measurements from quasar spectra as well as from Ly $\alpha$  emission from galaxies support a reionization scenario consistent with what we derive (Figure 20). The constraints from Ly $\alpha$  emission are heavily model dependent, and studies claiming a very low value of  $x_{\text{H II}}$  may be assuming a velocity offset of Ly $\alpha$  from systemic that is too high (e.g., Stark et al. 2014). While our results are in slight tension ( $1.3\sigma$ ) with the results from WMAP9, they are in excellent agreement with the more recent results from *Planck*. From our fiducial reionization history, we find  $\tau_{\text{es}} = 0.063 \pm 0.013$ , which is highly consistent with the measurement from *Planck* of  $\tau_{\text{es}} = 0.066 \pm 0.012$ . We remind the reader that our results did not use the *Planck* results as a constraint; rather, the inferred reionization history from our observed luminosity functions (with our assumptions on  $C$  and  $f_{\text{esc}}$ ) are in remarkable agreement with the *Planck* observations.

Future observations are necessary to improve the constraints on reionization from galaxies. Specifically, more robust measurements of the faint-end slope  $\alpha$  at  $z = 6\text{--}8$  can dramatically shrink the uncertainties on  $\rho_{\text{UV}}$ , subsequently reducing the width of our plausible values of  $x_{\text{H II}}$ . Likewise, improving the measurements at  $z \geq 9$  will inform us on whether the ionization fraction of the IGM at that early time was significantly non-zero. The HFF program will improve both of these areas, although definitive results will likely not be obtained until the *JWST* era.

## 9. EVOLUTION OF GALAXIES AT $z \geq 9$

Studies of galaxies at  $z \geq 9$  are now only in their nascent phase, but *HST* surveys such as CANDELS and UDF12 are beginning to probe this early epoch. The first robust results on galaxies in this epoch were published by Ellis et al. (2013), who used the new F140W data in the HUDF from the UDF12 program to discover galaxies at  $z \sim 9$ . This filter allows  $z \sim 9$  galaxies to be detected in two bands (F140W and F160W), dramatically reducing the contamination due to noise from F160W-only studies alone (Section 3.8.1; c.f., Bouwens et al. 2011a). Ellis et al. (2013) discovered the first robust sample at  $z > 8.5$ , finding seven candidate galaxies. McLure et al. (2013) followed this up with an analysis of the  $z = 9$  luminosity function, finding number densities at the faint end ( $M_{\text{UV}} \sim -18$ ) only slightly lower than at  $z = 8$ . Oesch et al. (2013) also analyzed the  $z = 9$  luminosity function, also finding seven  $z \sim 9$  candidate galaxies in the HUDF. Although the GOODS-S field lacks the F140W data necessary to detect potential  $z = 9$  galaxies in two bands, Oesch et al. (2013) added the full CANDELS/ERS GOODS-S field to improve their constraints at the bright end. However, they found no  $z = 9$  candidates in this larger field. Their published luminosity function is consistent with that from McLure et al. (2013) at the faint end. Bolstered by their additional constraints from the inclusion of non-detections from the larger GOODS-S field, Oesch et al. (2013) fit a luminosity function (keeping the faint-end slope and normalization fixed), finding a surprisingly faint value for  $M_{\text{UV}}^*$  of  $-18.8 \pm 0.3$ . However, this derivation was based on the assumption that the luminosity function shows luminosity evolution at  $z \geq 6$ —a trend that we have shown to be unlikely. Given this new insight, as well as the presence of a plethora of bright galaxies at  $z = 7$  and 8, we consider it likely that Oesch et al. (2013) underestimate the bright end of the  $z = 9$  luminosity function.

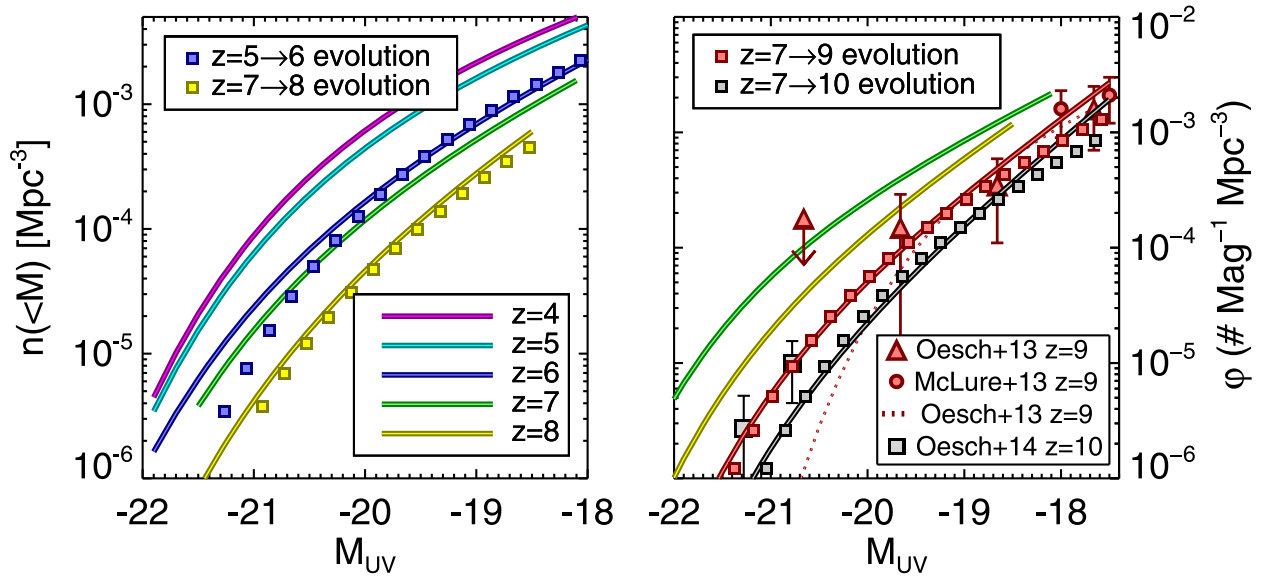
A number of recent papers have described empirical evidence that galaxies at high redshift have SFHs that increase with time (e.g., Finlator et al. 2011; Papovich et al. 2011; Jaacks et al. 2012b; Lundgren et al. 2014; Salmon et al. 2015). Most recently this has been examined by Salmon et al. (2015), who found that the SFRs of galaxies from  $z = 3$  to 6 are consistent with a power law of the form  $\Psi(t) = (t/\tau)^\gamma$  (with  $\gamma = 1.4 \pm 0.1$  and  $\tau = 92 \pm 14$  Myr). This analysis assumed that studying galaxies at a constant number density allows one to trace the progenitors and descendants of a galaxy population (e.g., van Dokkum et al. 2010; Leja et al. 2013), and their SFH was measured for a constant cumulative number density of  $2 \times 10^{-4} \text{ Mpc}^{-3}$ . Although the accuracy of this constant number density technique was initially studied at  $z < 3$ , recent evidence shows that it likely works out to  $z \sim 8$  (albeit with a possible slight evolution of number density with redshift; Behroozi et al. 2013; Jaacks et al. 2015).

Using our updated luminosity functions, we examine whether the SFHs at this earlier epoch are consistent with a similar functional form. Figure 21 shows the cumulative luminosity functions at  $z = 4$  to 8 from our analysis. Using the Salmon et al. rising SFH, we can evolve our  $z = 7$  cumulative luminosity function back in time to  $z = 8$  via:

$$\Psi_{z=8} = \Psi_{z=7} \left( \frac{t_{z=8}}{t_{z=7}} \right)^\gamma \quad (9)$$

where  $\Psi$  is the SFR,  $t_z$  is the cosmic time elapsed since formation to a given redshift, and using the Kennicutt (1998) conversion between  $M_{\text{UV}}$  and SFR (with the updated coefficient of 1.25). The available data cannot constrain the formation redshift ( $z_f$ ) because it is degenerate with the SFH exponent, so we assume a value of  $z_f = 20$ , which gives a close match between the predicted and observed  $z = 8$  cumulative luminosity functions. Figure 21 shows this predicted  $z = 8$  cumulative luminosity function alongside our observed one. A very close match is seen at nearly all magnitudes. Our predicted  $z = 8$  luminosity function slightly under-predicts the UV luminosity at  $M_{\text{UV}} > -19$ . However, as discussed above, our constraints on the faint end of the luminosity function at  $z = 8$  are tenuous at best. The agreement at the bright end is excellent. While we did not correct for dust in this analysis, dust is highly unlikely to change these results (particularly at the bright end where we are interested), because bright/massive galaxies at  $4 < z < 7$  all have similar UV slopes (Finkelstein et al. 2012b; Bouwens et al. 2014).

It is apparent when examining our cumulative luminosity functions in Figure 21 that this type of evolution will not work at all redshifts, as our luminosity functions are not uniformly spaced in magnitude (e.g., the  $z = 4$  and 5, and  $z = 6$  and 7 cumulative luminosity functions are very close together). We examined one other redshift, evolving the observed  $z = 5$  luminosity function to  $z = 6$ . We find a decent match, although it under-predicts the bright end and overpredicts the faint end. In any case, we are most interested in extrapolating to  $z > 8$ , so the fact that the predicted evolution works extremely well from  $z = 7$  to 8 gives us confidence that extrapolating to higher redshifts is reasonable. This assumed evolution is stronger than that observed from  $z = 6$  to  $z = 7$ . Had we assumed an SFH that matched the evolution from  $z = 6$  to  $z = 7$ , we would have over-predicted the  $z = 8$  LF. Our use of an SFH that matches



**Figure 21.** (Left) solid lines denote cumulative luminosity functions at  $z = 4$  to  $8$  from our study. We evolve the observed luminosity functions to higher redshift, assuming that the  $M_{UV} \propto \text{SFR}$ , that the SFR rises with time as  $\Psi \propto t^{1.4}$  (Salmon et al. 2015), and that galaxy progenitors and descendants share a common number density (e.g., Leja et al. 2013). We show two results: the predicted  $z = 6$  luminosity function evolved from  $z = 5$  (blue squares), and the predicted  $z = 8$  luminosity function evolved from  $z = 7$  (yellow squares). Though there are small discrepancies at  $z = 6$ , the match between predicted and observed is excellent at  $z = 8$ . (Right) our differential luminosity functions at  $z = 7$  and  $8$ , with  $z = 9$  data from the literature (triangles and circles). The red line shows our predicted  $z = 9$  ( $10$ ) luminosity function, continuing the evolution from  $z = 7$  as shown in the left panel. The red line shows our predicted  $z = 7$  Schechter function, with  $M_{UV}^* = -20.6$ . This predicted luminosity function shows excellent agreement with the observed faint end at  $z = 9$ , but is significantly higher at the bright end compared to the published luminosity function of Oesch et al. (2013). However, the recent discovery of bright  $z = 10$  candidate galaxies by Oesch et al. (2014; large gray squares) implies that bright galaxies are indeed present at this early epoch. We note that the  $z = 10$  number densities from Oesch et al. (2014) are actually more consistent with our  $z = 9$  predictions than  $z = 10$ ; clearly more work is needed to sort out the bright end at such early times. We conclude that when sufficient data exists for a large-volume survey for  $z = 9$  galaxies, large numbers of bright galaxies will be discovered.

the observed  $z = 7$  to  $z = 8$  evolution thus yields a conservatively low  $z = 9$  predicted luminosity function.

Given the relative paucity of observational information at  $z > 8$ , the fact that our assumed SFH matches the evolution from  $z = 7$  to  $8$  makes it interesting to continue our study out to  $z = 9$  (continuing to evolve the  $z = 7$  luminosity function due to its smaller uncertainties compared to our observed  $z = 8$  luminosity function). The red squares in Figure 21 show the expected  $z = 9$  luminosity function from our model, alongside our observed luminosity functions at  $z = 7$  and  $8$ . We calculated the expected  $z = 9$  luminosity function by again taking the  $z = 7$  luminosity function and evolving it out to  $z = 9$ , assuming the SFH discussed above (and  $z_f = 18$  for all number densities/magnitudes). As shown in this figure, our predicted  $z = 9$  luminosity function is consistent at the  $\sim 1\sigma$  level with all published data points from McLure et al. (2013) and Oesch et al. (2013). The insignificant underprediction at the faint end is due to the fact that our analysis effectively keeps the faint-end slope fixed to the  $z = 7$  value, while in reality it may become steeper. Given that our assumed SFH was derived from brighter galaxies (Salmon et al. 2015), the slight underprediction at the faint end is not surprising.

Figure 21 shows a Schechter function that matches our predicted  $z = 9$  luminosity function at all but the faintest magnitudes. To derive this Schechter function we used the measured significant evolution in  $\varphi^*$  and  $\alpha$ , with redshift shown in Figure 11 to find  $\varphi_{z=9}^* = 4.3 \times 10^{-5}$  and  $\alpha_{z=9} = -2.50$ . With these two parameters, we find a reasonable agreement with our predicted  $z = 9$  luminosity function with  $M_{z=9}^* = -20.6$ . This implies a slight dimming in the characteristic magnitude at  $z = 9$ , although much less than

that implied by Oesch et al. (2013), and within the uncertainties of our observed  $z = 7$  and  $8$  values. Regardless of the Schechter parameterization, our predicted luminosity function shows a much higher number density of bright galaxies than that reported by Oesch et al. (2013), yet is still moderately consistent with the observed data points from both Oesch et al. (2013) and McLure et al. (2013) at the faint end. Using our predicted  $z = 9$  Schechter function, we would expect to see  $\sim 3$   $z = 9$  galaxies at  $M_{UV} < -20.3$  ( $H < 27$ ) in a GOODS-sized field. Based on Poisson statistics alone, this is in mild tension with the zero galaxies at these magnitudes reported by Oesch et al. (2013). We also show results from this analysis when evolving  $z = 7$  to  $z = 10$ , which predicts  $M_{z=10}^* = -20.5$  when using our assumed evolution of  $\varphi^*$  and  $\alpha$  with redshift (or  $\sim 1$   $M_{UV} < -20.5$  galaxy per GOODS field).

Recently, Oesch et al. (2014) performed a new search for extremely distant galaxies, finding four bright  $z = 10$  candidates in the GOODS-N field and two new candidates from a re-analysis of the GOODS-S data set. These fields do not have deep F140W data, but Oesch et al. (2014) used the  $<3\sigma$  detections of these galaxies in the extremely shallow 800 s 3D-HST (PI van Dokkum) F140W pre-imaging data to place these galaxies at  $z = 10$ . Even though these galaxies are only detected in one band with *HST*, three are detected in IRAC (although at least one may be affected by blending from a nearby bright sources), thus their presence is intriguing. Figure 21 shows the number densities of these sources from Oesch et al. (2014); there is excellent agreement with our predicted  $z = 9$  evolution, although these data are much higher in abundance than our predicted  $z = 10$  luminosity function.



While these sources may be at  $z = 10$  rather than  $z = 9$ , if real, their presence confirms that bright galaxies are relatively abundant at  $z > 8.5$ .

Finally, we examine the change in the integrated luminosity density at  $z = 9$  with our proposed luminosity function compared to that from Oesch et al. (2013). Here we use our derived  $z = 9$  Schechter function, which matches our predictions at bright magnitudes, and matches the observed faint galaxy counts from the HUDF. The integrated luminosity density (from  $-23$  to  $-17$ ) is  $\sim 30\%$  higher than that published in Oesch et al. (2013). Thus, the precipitous decline in the luminosity density (Oesch et al. 2013, 2014) may be less than previously thought (e.g., Coe et al. 2013; Ellis et al. 2013; Behroozi & Silk 2015). While these results are intriguing, we conclude that in order to robustly probe the bright end of the  $z = 9$  luminosity function, we require a significantly increased searchable area with the correct filter set (allowing more than single-band detections) to discover these distant galaxies. Constraints on the full shape of the luminosity function in this distant epoch are crucial to design the most efficient surveys with *JWST*.

## 10. CONCLUSIONS

Combining the extremely deep data available in the HUDF with the still deep yet much wider data available from CANDELS in the GOODS-South and GOODS-North fields allows robust samples of galaxies to be discovered across a large dynamic range of UV luminosity at  $z = 4, 5, 6, 7$ , and  $8$ . Using a robust photometric redshift selection technique, we discovered a sample of nearly 7500 galaxies at  $3.5 < z < 8.5$  over five orders of UV magnitude, and over a volume of  $0.6\text{--}1.2 \times 10^6 \text{ Mpc}^3$ . We discovered a large number of bright ( $M_{\text{UV}} < -21$ ) galaxies at these redshifts, in excess of predictions based on previous estimates of the luminosity functions at  $z \geq 6$ .

1. Our sample selection performs very well when comparing to available spectroscopic redshifts. We perform various tests to estimate the contamination rate, which we find at worst to be  $\leq 15\%$ , and more likely to be  $\leq 5\text{--}10\%$ . This is consistent with contamination estimates based on the colors of the most likely contaminants, dusty star-forming galaxies at  $z \sim 2$ . Although the GOODS fields are only two of five CANDELS fields, the remaining three fields contain relatively shallow *Y*-band data, which can result in increased sample contamination, as well as a reduced ability to separate galaxies into  $z = 6, 7$ , and  $8$  samples.
2. Our large volume probed allows us to make a robust determination of the amplitude and shape of the bright end of the luminosity function, which can be used as a crucial probe of the physics dominating galaxy evolution. We used an MCMC technique to estimate the luminosity function and better characterize the uncertainties, both on the stepwise luminosity function and the parameters of the Schechter functional form. Our results agree with previous studies at the faint end, but deviate from some previous studies at the bright end, where our data allow us to better constrain the abundance of rare, bright galaxies. We find results consistent with a non-evolving characteristic magnitude ( $M_{\text{UV}}^* \approx -21$ ), with our values

of  $M_{\text{UV}}^*$  at  $z = 6$  and  $7$  brighter than the previous values of Bouwens et al. (2007, 2011b) at  $\sim 2\sigma$  significance. Both the faint-end slope ( $\alpha$ ) and the normalization ( $\varphi^*$ ) do significantly evolve with increasing redshift, to steeper and lower values, respectively. This is in contrast to previous results, which determined that the evolution of the luminosity function was primarily in luminosity.

3. We explored whether a Schechter functional form is required by the data, or whether a single (or double) power law is a better fit for our luminosity functions; a single power law form of the luminosity function may be expected at very high redshift, when dust may not be present, and/or feedback due to AGN activity is no longer sufficient to suppress star formation in the most massive galaxies. At  $z = 6$  and  $7$ , a Schechter function (or double power law) is required to fit the bright end. However, at  $z = 8$ , a single power law provides an equally good fit to the data. Although larger volumes will need to be probed to improve the estimates of the abundances of bright  $z = 8$  galaxies, if a power law is preferred, it could imply that we may be observing the era when feedback stops affecting massive galaxies. Comparing to semi-analytical models, we find that the evolution in our luminosity function can be explained by the changing impact of dust attenuation with redshift. In a future work we will explore whether this is a unique constraint, or whether a combination of feedback and dust changes can reproduce the observations.
4. We measure the evolution of the cosmic SFR density by integrating our observed luminosity functions to the observational limit of  $M_{\text{UV}} = -17$ , and correcting for dust attenuation. The cosmic SFR density evolves as  $(1+z)^{-4.3 \pm 0.5}$  at  $z \geq 4$ . This smoothly declining function with increasing redshift is consistent with published estimates of the SFR density at  $z \geq 9$ .
5. We investigate the constraints on the contribution of galaxies to reionization by integrating our luminosity functions down to  $M_{\text{UV}} = -13$ . Our fiducial results (assuming  $C/f_{\text{esc}} = 23$ , which does not violate Ly $\alpha$  forest constraints at  $z \leq 6$ ) are consistent with a reionization history that begins at  $z > 10$ , and completes at  $z \approx 6$ , with a midpoint at  $6.7 < z < 9.4$ . However, the uncertainties particularly at  $z \geq 7$  are high, due to the relatively high uncertainty in the faint-end slope, such that our observations are consistent with an IGM at  $z = 8$  that is anywhere from completely ionized to 85% neutral.
6. The presence of bright galaxies at  $z = 6\text{--}8$  has interesting implications for the luminosity functions at higher redshift. We used empirically derived SFHs to evolve our  $z = 7$  luminosity function back to  $z = 9$ , and predict that  $\sim 3$  bright ( $M_{\text{UV}} < -20.3$ ) galaxies should be detectable per GOODS-sized field. This is contrary to initial observational results, which, using single-band detections, found no bright  $z = 9$  galaxies, but consistent with emerging results that some bright galaxies may exist at  $z = 10$ . Future wider-area studies with two-band detections will provide a more robust estimate of the bright end of the  $z = 9$  luminosity function.

This study highlights the power of combining deep and wide-area studies to probe galaxy populations at very high redshifts, a topic that will remain highly active through the advent of *JWST*. These results leave us with a variety of

questions. What is responsible for the apparent abundance of bright galaxies at  $z > 6$ ? Is this tied in with a reduction of feedback, or is some other physical process in play? Does this trend continue out to higher redshifts, or does the luminosity density fall off dramatically at  $z > 8$ , as has been proposed? Although these issues are inherently intertwined, we can make progress on them with future wide-area *HST* surveys. This will provide the most complete view of the high-redshift universe by the end of this decade, allowing us to make full use of *JWST*.

We thank Kristian Finlator, Brian Siana, Rychard Bouwens, Pascal Oesch, Dan Jaffe, and Jon Trump for useful conversations. S.L.F. acknowledges support from the University of Texas at Austin College of Natural Sciences. M.S. was supported by a NASA Astrophysics and Data Analysis Program award issued by JPL/Caltech. R.J.M. acknowledges ERC funding via the award of a consolidator grant. This work is based on observations made with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with program #12060. This work is also based in part on observations made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech.

## REFERENCES

- Alavi, A., Siana, B., Richard, J., et al. 2014, *ApJ*, **780**, 143  
 Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, *AJ*, **126**, 539  
 Anderson, J., & Bedin, L. R. 2010, *PASP*, **122**, 1035  
 Arnouts, S., Schiminovich, D., Ilbert, O., et al. 2005, *ApJL*, **619**, L43  
 Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, *ApJ*, **769**, 80  
 Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, *ApJS*, **218**, 33  
 Atek, H., Siana, B., Scarlata, C., et al. 2011, *ApJ*, **743**, 121  
 Balestra, I., Mainieri, V., Popesso, P., et al. 2010, *A&A*, **512**, A12  
 Barger, A. J., Cowie, L. L., & Wang, W.-H. 2008, *ApJ*, **689**, 687  
 Barone-Nugent, R. L., Trenti, M., Wyithe, J. S. B., et al. 2014, *ApJ*, **793**, 17  
 Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, *MNRAS*, **447**, 3402  
 Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, **132**, 1729  
 Behroozi, P. S., Marchesini, D., Wechsler, R. H., et al. 2013, *ApJL*, **777**, L10  
 Behroozi, P. S., & Silk, J. 2015, *ApJ*, **799**, 32  
 Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393  
 Blanc, G. A., Adams, J. J., Gebhardt, K., et al. 2011, *ApJ*, **736**, 31  
 Bolton, J. S., & Haehnelt, M. G. 2007, *MNRAS*, **382**, 325  
 Bolton, J. S., Haehnelt, M. G., Warren, S. J., et al. 2011, *MNRAS*, **416**, L70  
 Bouwens, R. J., Illingworth, G. D., Blakeslee, J. P., & Franx, M. 2006, *ApJ*, **653**, 53  
 Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, *ApJ*, **670**, 928  
 Bouwens, R. J., Illingworth, G. D., Labbe, I., et al. 2011a, *Natur*, **469**, 504  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2010a, *ApJL*, **709**, L133  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2010b, *ApJL*, **708**, L69  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011b, *ApJ*, **737**, 90  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, *ApJ*, **754**, 83  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, *ApJ*, **793**, 115  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, **803**, 34  
 Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2012, *MNRAS*, **426**, 2772  
 Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2014, *MNRAS*, **440**, 2810  
 Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2015, *MNRAS*, **452**, 1817  
 Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503  
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000  
 Bunker, A. J., Stanway, E. R., Ellis, R. S., McMahon, R. G., & McCarthy, P. J. 2003, *MNRAS*, **342**, L47  
 Bunker, A. J., Wilkins, S., Ellis, R. S., et al. 2010, *MNRAS*, **409**, 855  
 Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682  
 Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, *ApJ*, **429**, 582  
 Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, **345**, 245  
 Casey, C. M., Narayanan, D., & Cooray, A. 2014a, *PhR*, **541**, 45  
 Casey, C. M., Scoville, N. Z., Sanders, D. B., et al. 2014b, *ApJ*, **796**, 95  
 Cash, W. M. 1979, *ApJ*, **228**, 939  
 Castellano, M., Fontana, A., Boutsia, K., et al. 2010, *A&A*, **511**, A20  
 Coe, D., Zitrin, A., Carrasco, M., et al. 2013, *ApJ*, **762**, 32  
 Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, *AJ*, **112**, 839  
 Cucciati, O., Tresse, L., Ilbert, O., et al. 2012, *A&A*, **539**, A31  
 Curtis-Lake, E., McLure, R. J., Dunlop, J. S., et al. 2014, arXiv:1409.1832  
 Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, *ApJ*, **775**, 93  
 Dayal, P., Dunlop, J. S., Maio, U., & Ciardi, B. 2013, *MNRAS*, **434**, 1486  
 Dickinson, M., Hanley, C., Elston, R., et al. 2000, *ApJ*, **531**, 624  
 Dickinson, M., Stern, D., Giavalisco, M., et al. 2004, *ApJL*, **600**, L99  
 Dunlop, J. S., McLure, R. J., Robertson, B. E., et al. 2012, *MNRAS*, **420**, 901  
 Dunlop, J. S., Rogers, A. B., McLure, R. J., et al. 2013, *MNRAS*, **432**, 3520  
 Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, *ApJL*, **763**, L7  
 Faisst, A. L., Capak, P., Carollo, C. M., Scarlata, C., & Scoville, N. 2014, *ApJ*, **788**, 87  
 Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, **132**, 117  
 Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, *ApJ*, **688**, 85  
 Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, *ApJS*, **154**, 10  
 Ferguson, H. C., Dickinson, M., Giavalisco, M., et al. 2004, *ApJL*, **600**, L107  
 Finkelstein, S. L., Hill, G. J., Gebhardt, K., et al. 2011, *ApJ*, **729**, 140  
 Finkelstein, S. L., Papovich, C., Dickinson, M., et al. 2013, *Natur*, **502**, 524  
 Finkelstein, S. L., Papovich, C., Giavalisco, M., et al. 2010, *ApJ*, **719**, 1250  
 Finkelstein, S. L., Papovich, C., Ryan, R. E., et al. 2012a, *ApJ*, **758**, 93  
 Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012b, *ApJ*, **756**, 164  
 Finlator, K., Oh, S. P., Özel, F., & Davé, R. 2012, *MNRAS*, **427**, 2464  
 Finlator, K., Oppenheimer, B. D., & Davé, R. 2011, *MNRAS*, **410**, 1703  
 Fioc, M., & Rocca-Volmerange, B. 1997, *A&A*, **326**, 950  
 Galametz, A., Grazian, A., Fontana, A., et al. 2013, *ApJS*, **206**, 10  
 Gehrels, N. 1986, *ApJ*, **303**, 336  
 Giavalisco, M., Dickinson, M., Ferguson, H. C., et al. 2004, *ApJL*, **600**, L103  
 Grazian, A., Castellano, M., Fontana, A., et al. 2012, *A&A*, **547**, A51  
 Grazian, A., Fontana, A., de Santis, C., et al. 2006, *A&A*, **449**, 951  
 Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, **197**, 35  
 Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *ApJS*, **207**, 24  
 Hathi, N. P., Malhotra, S., & Rhoads, J. E. 2008, *ApJ*, **673**, 686  
 Hathi, N. P., Mobasher, B., Capak, P., Wang, W.-H., & Ferguson, H. C. 2012, *ApJ*, **757**, 43  
 Hayes, M., Schaerer, D., Östlin, G., et al. 2011, *ApJ*, **730**, 8  
 Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, **208**, 19  
 Inoue, A. K. 2011, *MNRAS*, **415**, 2920  
 Iwata, I., Inoue, A. K., Matsuda, Y., et al. 2009, *ApJ*, **692**, 1287  
 Jaacks, J., Choi, J.-H., Nagamine, K., Thompson, R., & Varghese, S. 2012a, *MNRAS*, **420**, 1606  
 Jaacks, J., Finkelstein, S. L., & Nagamine, K. 2015, *ApJ*, submitted (arXiv:1507.00713)  
 Jaacks, J., Nagamine, K., & Choi, J. H. 2012b, *MNRAS*, **427**, 403  
 Kennicutt, R. C., Jr. 1998, *ARA&A*, **36**, 189  
 Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, **197**, 36  
 Koekemoer, A. M., Ellis, R. S., McLure, R. J., et al. 2013, *ApJS*, **209**, 3  
 Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, *ApJS*, **192**, 18  
 Kriek, M., Labbé, I., Conroy, C., et al. 2010, *ApJL*, **722**, L64  
 Kuhlen, M., & Faucher-Giguère, C.-A. 2012, *MNRAS*, **423**, 862  
 Kurk, J., Cimatti, A., Daddi, E., et al. 2013, *A&A*, **549**, A63  
 Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, *ApJL*, **777**, L19  
 Laidler, V. G. 2007, *PASP*, **119**, 1325  
 Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, **379**, 1599  
 Leja, J., van Dokkum, P., & Franx, M. 2013, *ApJ*, **766**, 33  
 Liddle, A. R. 2004, *MNRAS*, **351**, L49  
 Lorenzoni, S., Bunker, A. J., Wilkins, S. M., et al. 2013, *MNRAS*, **429**, 150  
 Lundgren, B. F., van Dokkum, P., Franx, M., et al. 2014, *ApJ*, **780**, 34  
 Madau, P. 1995, *ApJ*, **441**, 18  
 Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415  
 Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, *MNRAS*, **283**, 1388  
 Madau, P., Haardt, F., & Rees, M. J. 1999, *ApJ*, **514**, 648  
 McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, **544**, A156  
 McLure, R. J., Cirasuolo, M., Dunlop, J. S., Foucaud, S., & Almaini, O. 2009, *MNRAS*, **395**, 2196  
 McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, *MNRAS*, **432**, 2696  
 McLure, R. J., Dunlop, J. S., Cirasuolo, M., et al. 2010, *MNRAS*, **403**, 960  
 McQuinn, M., Oh, S. P., & Faucher-Giguère, C.-A. 2011, *ApJ*, **743**, 82

- Merlin, E., Fontana, A., Ferguson, H. C., et al. 2015, *A&A*, submitted (arXiv:1505.02516)
- Mesinger, A. 2010, *MNRAS*, 407, 1328
- Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, *Natur*, 474, 616
- Mostardi, R. E., Shapley, A. E., Nestor, D. B., et al. 2013, *ApJ*, 779, 65
- Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H.-W. 2011, *ApJ*, 731, 113
- Nestor, D. B., Shapley, A. E., Kornei, K. A., Steidel, C. C., & Siana, B. 2013, *ApJ*, 765, 47
- Nestor, D. B., Shapley, A. E., Steidel, C. C., & Siana, B. 2011, *ApJ*, 736, 18
- Newman, J. A., & Davis, M. 2002, *ApJ*, 564, 567
- Oesch, P. A., Bouwens, R. J., Carollo, C. M., et al. 2010a, *ApJL*, 709, L21
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2010b, *ApJL*, 709, L16
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2012, *ApJ*, 759, 135
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2013, *ApJ*, 773, 75
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2014, *ApJ*, 786, 108
- Oesch, P. A., Stiavelli, M., Carollo, C. M., et al. 2007, *ApJ*, 671, 1212
- Oke, J. B., & Gunn, J. E. 1983, *ApJ*, 266, 713
- Ono, Y., Ouchi, M., Curtis-Lake, E., et al. 2013, *ApJ*, 777, 155
- Ono, Y., Ouchi, M., Mobasher, B., et al. 2012, *ApJ*, 744, 83
- Ouchi, M., Mobasher, B., Shimasaku, K., et al. 2009, *ApJ*, 706, 1136
- Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, *ApJ*, 723, 869
- Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Giavalisco, M. 2011, *MNRAS*, 412, 1123
- Pawlik, A. H., Schaye, J., & van Scherpenzeel, E. 2009, *MNRAS*, 394, 1812
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *AJ*, 124, 266
- Pentericci, L., Fontana, A., Vanzella, E., et al. 2011, *ApJ*, 743, 132
- Pentericci, L., Vanzella, E., Fontana, A., et al. 2014, *ApJ*, 793, 113
- Planck Collaboration Ade, P. A. R., Aghanim, N., et al. 2015, arXiv:1502.01589
- Reddy, N. A., & Steidel, C. C. 2009, *ApJ*, 692, 778
- Rhoads, J. E., Malhotra, S., Pirzkal, N., et al. 2009, *ApJ*, 697, 942
- Rhoads, J. E., Malhotra, S., Stern, D., et al. 2013, *ApJ*, 773, 32
- Robertson, B. E., Ellis, R. S., Dunlop, J. S., et al. 2014, *ApJL*, 796, L27
- Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, *ApJ*, 768, 71
- Ryan, R. E., Thorman, P. A., Yan, H., et al. 2011, *ApJ*, 739, 83
- Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, 799, 183
- Salvaterra, R., Ferrara, A., & Dayal, P. 2011, *MNRAS*, 414, 847
- Sanders, D. B., Mazzarella, J. M., Kim, D.-C., Surace, J. A., & Soifer, B. T. 2003, *AJ*, 126, 1607
- Schechter, P. 1976, *ApJ*, 203, 297
- Schenker, M. A., Ellis, R. S., Konidakis, N. P., & Stark, D. P. 2014, *ApJ*, 795, 20
- Schenker, M. A., Robertson, B. E., Ellis, R. S., et al. 2013, *ApJ*, 768, 196
- Schimminovich, D., Ilbert, O., Arnouts, S., et al. 2005, *ApJL*, 619, L47
- Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
- Schmidt, K. B., Treu, T., Trenti, M., et al. 2014, *ApJ*, 786, 57
- Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., & Erb, D. K. 2006, *ApJ*, 651, 688
- Siana, B., Teplitz, H. I., Ferguson, H. C., et al. 2010, *ApJ*, 723, 241
- Smit, R., Bouwens, R. J., Labbé, I., et al. 2014, *ApJ*, 784, 58
- Somerville, R. S., Gilmore, R. C., Primack, J. R., & Domínguez, A. 2012, *MNRAS*, 423, 1992
- Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, *MNRAS*, 391, 481
- Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2015, *ApJ*, submitted (arXiv:1507.05636)
- Stanway, E. R., McMahon, R. G., & Bunker, A. J. 2005, *MNRAS*, 359, 1184
- Stark, D. P., Richard, J., Charlot, S., et al. 2014, *MNRAS*, 445, 3200
- Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, *ApJ*, 763, 129
- Steidel, C. C., Pettini, M., & Adelberger, K. L. 2001, *ApJ*, 546, 665
- Szokoly, G. P., Bergeron, J., Hasinger, G., et al. 2004, *ApJS*, 155, 271
- Tilvi, V., Papovich, C., Finkelstein, S. L., et al. 2014, *ApJ*, 794, 5
- Tilvi, V., Papovich, C., Tran, K.-V. H., et al. 2013, *ApJ*, 768, 56
- Trenti, M., Bradley, L. D., Stiavelli, M., et al. 2011, *ApJL*, 727, L39
- Trenti, M., Perna, R., Levesque, E. M., Shull, J. M., & Stocke, J. T. 2012, *ApJL*, 749, L38
- van der Burg, R. F. J., Hildebrandt, H., & Erben, T. 2010, *A&A*, 523, A74
- van der Wel, A., Straughn, A. N., Rix, H.-W., et al. 2011, *ApJ*, 742, 111
- van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, *ApJ*, 709, 1018
- Vanzella, E., Cristiani, S., Dickinson, M., et al. 2008, *A&A*, 478, 83
- Vanzella, E., Giavalisco, M., Dickinson, M., et al. 2009, *ApJ*, 695, 1163
- Vanzella, E., Giavalisco, M., Inoue, A. K., et al. 2010, *ApJ*, 725, 1011
- Vanzella, E., Guo, Y., Giavalisco, M., et al. 2012, *ApJ*, 751, 70
- Wilkins, S. M., Bunker, A. J., Stanway, E., Lorenzoni, S., & Caruana, J. 2011, *MNRAS*, 417, 717
- Willott, C. J., McLure, R. J., Hiben, P., et al. 2013, *AJ*, 145, 4
- Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, *ApJS*, 193, 27
- Wuyts, S., van Dokkum, P. G., Franx, M., et al. 2009, *ApJ*, 706, 885
- Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, *ApJS*, 195, 10
- Yan, H., Finkelstein, S. L., Huang, K.-H., et al. 2012, *ApJ*, 761, 177
- Yan, H., Yan, L., Zamojski, M. A., et al. 2011, *ApJL*, 728, L22